Multivariate Spatial Process Gradients with Environmental Applications

by

Maria A. Terres

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
Alan E. Gelfand, Supervisor

_____
Robert L. Wolpert

_____
Mine Çetinkaya-Rundel

_____
James S. Clark

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University
2014

Abstract

Multivariate Spatial Process Gradients with Environmental Applications

by

Maria A. Terres

Department of Statistical Science
Duke University

Date: _____

Approved:

_____
Alan E. Gelfand, Supervisor

_____
Robert L. Wolpert

_____
Mine Çetinkaya-Rundel

_____
James S. Clark

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistical Science in the Graduate School of Duke University
2014

Abstract

Previous papers have elaborated formal gradient analysis for spatial processes, focusing on the distribution theory for directional derivatives associated with a response variable assumed to follow a Gaussian process model. In the current work, these ideas are extended to additionally accommodate one or more continuous covariate(s) whose directional derivatives are of interest and to relate the behavior of the directional derivatives of the response surface to those of the covariate surface(s). It is of interest to assess whether, in some sense, the gradients of the response follow those of the explanatory variable(s), thereby gaining insight into the local relationships between the variables. The joint Gaussian structure of the spatial random effects and associated directional derivatives allows for explicit distribution theory and, hence, kriging across the spatial region using multivariate normal theory. The gradient analysis is illustrated for bivariate and multivariate spatial models, non-Gaussian responses such as presence-absence and point patterns, and outlined for several additional spatial modeling frameworks that commonly arise in the literature. Working within a hierarchical modeling framework, posterior samples enable all gradient analyses to occur as post model fitting procedures.

Dedicated to all of the formal and informal teachers I've had along the way.

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Symbols

$\mathbb{R}^d$	Set of all d-dimensional real-valued vectors.
$N(m, v)$	Multivariate normal with mean m and covariance matrix v.
$E(x)$	Expected value of x.
$Var(x)$	Variance of x.
$SD(x)$	Standard deviation of x.
$Cov(x, y)$	Covariance of x and y.
$\Phi(x)$	Standard normal cdf evaluated at x.
$\phi(x)$	Standard normal pdf evaluated at x.

Abbreviations

CAR	Conditionally autogregressive.
CDF	Cumulative distribution function.
CFR	Cape Floristic Region.
FWC	Leaf fresh water content.
GLM	Generalized linear model.
GP	Gaussian process.
GPS	Global positioning system.
GWR	Geographically weighted regression.
LMA	leaf dry mass per unit area.

MAP	Mean annual precipitation.
MAT	Mean annual temperature.
MCMC	Markov chain Monte carlo.
PDF	Probability density function.
SVCP	Spatially varying coefficient process.

Acknowledgements

I would like to thank my advisor, Alan Gelfand, for his encouragement and support over the past few years. Our many discussions spanning past, present and future have undoubtedly shaped my career and life as a statistical researcher. I am grateful to Jim Clark and the other members of the CDI group for engaging discussions each week. I am appreciative of Robert Wolpert who was always available to hear my joys and concerns, as well as Mine Çetinkaya-Rundel, Kari Lock-Morgan and Dalene Stangl for supporting my summer teaching endeavors. Finally, a big thank you to my friends and family for believing in me.

In addition, I would like to acknowledge National Science Foundation (NSF), International Society for Bayesian Analysis (ISBA), American Statistical Association (ASA), Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Duke University for providing financial support of my education.

1

Introduction

1.1 Geostatistical Modeling and Gaussian Processes

1.1.1 The Basics of Geostatistical Modeling

Due in part to advances in various technologies, such as Global Positioning Systems (GPS), recording geocoded locations has become a routine aspect of data collection. Such data allow for exploratory and statistical investigation into how the outcome of interest behaves across different spatial regions. While covariate information may explain a substantial portion of the variation in response, there is often underlying spatial structure that is difficult to account for and may bias results if ignored. To accommodate the abundance of geo-referenced data, spatial process models have become widespread and are outlined in detail in several articles and books (e.g. Banerjee et al., 2004; Cressie and Wikle, 2011). While spatial modeling is appropriate in a variety of contexts, the examples and applications explored throughout this text will be focused on ecology and the environment.

For a region of interest D, the set of conceptual responses $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ can be viewed as a realization of a random surface observed at a finite set of locations. The

general model for such data is of the form,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{1.1}$$

where $\mu(\mathbf{s})$ is a location-specific mean generally modeled as a linear function of some covariates, $\epsilon(\mathbf{s})$ is a pure error term modeled as a zero-mean white noise process, and $w(\mathbf{s})$ is a spatial error term modeled as a zero-mean stationary Gaussian process. For some data, such as elevation, the spatial-error term may be sufficient and $\epsilon(\mathbf{s})$ will be removed from the model.

A spatial process with density $p(\cdot)$ is defined over any set of locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ such that $p(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ is consistent across permutations and marginalizations. If the process is Gaussian, then the associated density will be multivariate normal and characterized by a covariance function, $c(\cdot)$. A valid covariance function must be a positive definite function, and may exhibit additional properties such as stationarity and isotropy. A process is strictly stationary when $p(\mathbf{s}_1, \ldots, \mathbf{s}_n) = p(\mathbf{s}_1+\boldsymbol{\delta}, \ldots, \mathbf{s}_n+\boldsymbol{\delta})$ for some separation vector $\boldsymbol{\delta}$, and it is weakly stationary when $\mu(\mathbf{s}) = \mu$ for all $\mathbf{s}$ and $Cov(Y(\mathbf{s}), Y(\mathbf{s}+\boldsymbol{\delta})) = c(\boldsymbol{\delta})$. A Gaussian process is weakly stationary if and only if it is strictly stationary, so throughout this text we will refer to processes as simply "stationary". A stationary covariance function is isotropic if it depends only on the length of the separation vector $\boldsymbol{\delta}$, $c(\boldsymbol{\delta}) = c(||\boldsymbol{\delta}||)$.

The Gaussian processes illustrated in the following chapters will typically be mean-zero with stationary isotropic covariance functions. In particular we focus on the flexible class of Matérn covariance functions with mean square differentiability of process realizations controlled by a smoothness parameter ν. In our examples it is desirable to have once, but not twice, mean square differentiability, so the Matérn covariance with $\nu = 3/2$ is selected: $c(||\boldsymbol{\delta}||) = \sigma^2(1 + \phi||\boldsymbol{\delta}||)\exp(-\phi||\boldsymbol{\delta}||)$.

A hierarchical Bayesian approach is used to fit the model in (1.1). Letting $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))'$ represent the observed responses, $\mathbf{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_n))'$

represent the associated spatial random effects, and X be the matrix of covariates, the model can be written

$$\begin{aligned} \mathbf{Y}|\mathbf{W}, \boldsymbol{\theta} &\sim N(X\beta + \mathbf{W}, \tau^2) \\ \mathbf{W}|\boldsymbol{\theta} &\sim N(\mathbf{0}, K(\cdot|\sigma^2, \phi)) \\ p(\boldsymbol{\theta}) &= p(\beta)p(\tau^2)p(\sigma^2)p(\phi) \end{aligned} \tag{1.2}$$

where $K(\cdot|\sigma^2, \phi)$ is a covariance matrix with entry i, j evaluated at $\boldsymbol{\delta} = \mathbf{s}_i - \mathbf{s}_j$, and $p(\boldsymbol{\theta})$ is the prior on the model parameters. The model is fit using MCMC methods, and to improve fitting one would generally marginalize over the spatial random effects $\mathbf{W}$ when fitting this model. However, marginalization will not be possible when the first-stage is non-Gaussian; instead, the spatial random effects will need to be sampled at each iteration.

While data is only observed at a finite set of locations, e.g. pollution concentrations measured at monitoring stations, researchers are generally interested in how the data varies across the entire region including unobserved locations. One advantage of the spatial model described above is that it enables straight forward prediction at unobserved locations, also known as kriging. This interpolation takes place via the posterior predictive distribution. For an unobserved location $\mathbf{s}_0$,

$$p(Y(\mathbf{s}_0)|X(\mathbf{s}_0), \mathbf{Y}, \mathbf{X}) = \int p(Y(\mathbf{s}_0)|X(\mathbf{s}_0), \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) d\boldsymbol{\theta}$$

with samples $Y(\mathbf{s}_0)$ drawn from $p(Y(\mathbf{s}_0)|X(\mathbf{s}_0), \mathbf{Y}, \mathbf{X})$ via composition. For a Gaussian first-stage $p(Y(\mathbf{s}_0)|X(s_0), \mathbf{Y}, \boldsymbol{\theta})$ is easily computed using multivariate normal theory.

1.1.2 More Advanced Geostatistical Modeling

Spatial modeling goes well beyond the spatial linear model defined in (1.1), with modifications enabling application to a wide variety of problems and datasets. Examples of such extensions include multivariate spatial models, spatial models for

non-Gaussian responses, spatial-temporal models, spatially-varying coefficient models, and several adaptations to improve computational efficiency for large datasets. These topics are briefly reviewed here.

When there are several variables of interest whose surfaces can all reasonably be assumed to be spatially smooth their behavior can be captured through a joint spatial model. Several frameworks exist for multivariate spatial modeling including some simple separable models as well as more constructive non-separable approaches. In our demonstrations we focus on the coregionalization model which can either be built through successive conditioning of the variables or through specification of a joint dependence on latent Gaussian processes (Royle and Berliner, 1999; Banerjee et al., 2004). Within this framework there will be an equal number of latent Gaussian processes and spatial variables, although each spatial variable may be dependent on multiple Gaussian processes to induce dependence between the variables.

For some applications the response variable of interest cannot be assumed to follow a normal distribution. For example, presence-absence observations for a plant or animal species correspond to a binary response, making a binomial distribution a more appropriate first-stage in (1.2). In general, a spatial generalized linear model (GLM) can be fit for these types of non-Gaussian data (e.g. Heagerty and Lele, 1998; Paciorek, 2007; Berrett and Calder, 2012). These spatial GLMs utilize a link function to relate a latent Gaussian process and a continuous response, such as presence probability. As in the standard spatial process model, the latent Gaussian process encourages similar outcomes for proximate locations.

Many environmental quantities are available at a fine temporal scale, e.g. daily pollutant levels, indicating a need for combined spatial-temporal modeling. Similar to the multivariate scenario, these models may be separable treating space and time as independent, or may be non-separable. Generally the non-separable models will be preferable, since space and time are often expected to interact. Spatial-temporal

models have been laid out in several books, (e.g. Wikle et al., 1998; Banerjee et al., 2004; Cressie and Wikle, 2011), and have been adapted and extended in the literature, (e.g. Gelfand et al., 2005; Reich et al., 2011b). These models allow for examination of temporally-varying spatial relationships between variables.

To allow for spatially varying relationships between variables one can fit a spatially varying coefficient model. The most common approaches for such models are the spatially varying coefficient process (SVCP) proposed by Gelfand et al. (2003) and the geographically weighted regression (GWR) proposed by Fotheringham et al. (2002). The former is a Bayesian model in which regression coefficients are treated as spatially correlated processes, while the latter estimates parameters using spatial weights to emphasize proximate observations. Comparing the relative benefits of the two models, simulation studies suggest the coefficient estimates under SVCP are more accurate and more robust to collinearity than GWR and provides more formal measures of model uncertainty (Wheeler and Calder, 2007; Wheeler and Waller, 2009). However, both models can be computationally intensive, and make assumptions regarding the variables' relationship that go beyond a standard spatial regression model.

With the above as examples, there is clearly a wide array of spatial modeling techniques that accommodate many types of data and allow for flexible relationships between variables. However, many of these models suffer from long computation times when subjected to large datasets common for environmental applications. The primary computational expenses are the high-dimensional matrix decompositions related to the covariance matrix in the Gaussian process model. Recent approaches for reducing computation times generally favor working in a lower dimensional subspace to approximate the full likelihood (e.g. Wikle and Cressie, 1999; Higdon, 2002; Ver Hoef et al., 2004; Kim et al., 2005; Fuentes, 2007; Banerjee et al., 2008; Stein, 2013). Each approximation corresponds to a unique set of advantages and disadvan-

tages, but all share the common goal of speeding up computation time to accommodate large spatial data.

1.2 Spatial Modeling for Environmental Applications

Environmental data are increasingly collected across large spatial regions and are recorded with geo-referenced locations. Examples of environmental point-referenced data include pollutant levels, temperature, elevation and, when the spatial scale is large enough, plot-level ecological data such as species presence. In recent years the environmental literature has seen arguments in favor of Bayesian hierarchical modeling for both spatial and non-spatial data (Clark, 2005; Clark and Gelfand, 2006; Cressie et al., 2009). The complex systems within which environmental and ecological variables interact make the Bayesian treatment of uncertainty an appealing and realistic framework. Similarly, accounting for unmeasured spatial variation can improve model inference and allow for interpolation across the region. Examples of spatial models for environmental applications are abundant in the literature including capture-recapture data (Royle et al., 2013), pollution levels (Reich et al., 2011a), tropical tree growth (Baribault et al., 2012), among many others.

1.3 Spatial Gradient Processes

Spatial gradients under Gaussian processes were elaborated in Banerjee et al. (2003) to address the rate of change of a spatial surface at a given point in a given direction. Their paper defines directional derivative processes with corresponding distribution theory to enable interpolation across a region. For a mean square differentiable process $Y(\mathbf{s})$, they define the directional derivative in the direction $\mathbf{u}$ as

$$D_{\mathbf{u}}Y(\mathbf{s}) = \lim_{h \to 0} \frac{Y(\mathbf{s} + h\mathbf{u}) - Y(\mathbf{s})}{h} \tag{1.3}$$

For a spatial process defined in $\mathbb{R}^2$, the gradient vector is defined in the directions forming an orthonormal basis: $\nabla_Y(\mathbf{s}) = (D_{\mathbf{e_1}}Y(\mathbf{s}), D_{\mathbf{e_2}}Y(\mathbf{s}))'$ where $\mathbf{e_i}$ is a 2×1 vector of 0s with a 1 in the i^{th} entry. For any direction of interest $\mathbf{u}$, the directional derivative can be computed using the gradient vector, $D_{\mathbf{u}}Y(\mathbf{s}) = \mathbf{u}'\nabla_Y(\mathbf{s})$. Jointly with the parent $Y(\mathbf{s})$ process, the gradient vectors follow a multivariate Gaussian process with distribution fully determined by parameters in the parent spatial model, $[Y(\mathbf{s})|\boldsymbol{\theta}]$. Thus, computing the posterior predictive distribution for $\nabla_Y(\mathbf{s})|Y(\mathbf{s})$ is simply an application of multivariate normal theory, and posterior draws of the gradient vectors can occur via composition using the posterior samples of the parameters for the parent model. As such, all gradient analysis occurs post model fitting.

Subsequent papers addressed the smoothness of the gradient processes (Banerjee and Gelfand, 2003), applied the methodology to models with spatially varying coefficients (Majumdar et al., 2006), extended the methods to allow for curvilinear or boundary analysis (Banerjee and Gelfand, 2006), and explored analysis for non-Gaussian spatial processes, in particular, the spatial Dirichlet processes (Guindani and Gelfand, 2006). All of the current literature on spatial gradient processes focuses on the rates of change of a response surface, with the mean surface modeled as a linear function of a set of fixed covariates.

Spatial gradient processes can be helpful in illuminating environmental patterns across spatial regions. For example, ecologists are interested in characterizing how plants' range boundaries relate to climate (e.g. Canham and Thomas, 2010; Thuiller et al., 2004; Thomas, 2010). One aspect of this may be to consider the rate of change for plant frequencies along a temperature gradient. These kinds of studies can provide insight into potential effects of climate change on plants' range distributions.

1.4 Thesis Contribution and Outline

Consider a response-covariate pair measured within a spatial region. Particularly within the complex context of environmental and ecological data, it is unlikely that the covariate will have a uniform effect on the response variable across the entire region. Instead there may be pockets where the covariate has a negative influence, pockets where it has a positive influence, and/or pockets where it has a negligible influence. Possible explanations for such variation in the relationship include spatial variation in unaccounted for covariates, the magnitude of the response or covariate, or the rate of change of the response or covariate. Examination of how this relationship is varying spatially is a desirable alternative to painstakingly trying to model every possible explanatory factor.

One approach to understanding spatially varying relationships is to allow for spatially varying coefficients as in the GWR and SVCP models discussed in subsection 1.1.2. However, these models can be complicated to fit and rely on a number of underlying assumptions. Based on ideas from sensitivity analysis we propose an alternative methodology that allows for fitting a less complicated model to the data, followed by a post model fitting procedure to examine spatial variation in the variable relationship(s).

Given a relationship between two quantities, sensitivity analysis compares the relative rates of change using ratios of derivatives (see e.g. Tomovic and Vukobratovic, 1972). This analysis allows one to study how each quantity reacts to changes in the other. This is directly analogous to the goals when fitting a regression. For example, consider the spatial linear regression

$$Y(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w(\mathbf{s}) \tag{1.4}$$

where $w(\mathbf{s})$ is a mean-zero Gaussian process and $X(\mathbf{s})$ is some covariate. The parameter β_1 informs about the change in expected value of $Y(\mathbf{s})$ relative to a given change

in $X(\mathbf{s})$ and can be interpreted as a ratio of derivatives, $\beta_1 = dE(Y(\mathbf{s}))/dX(\mathbf{s})$. This parameter is constant across the region and can be said to define the "global" relationship between $Y(\mathbf{s})$ and $X(\mathbf{s})$.

Although the β_1 parameter in (1.4) is not spatially varying and cannot inform on the spatial relationship, it takes a small leap to consider replacing the derivatives in the above ratio with spatially varying derivatives. I.e., to compare the relative rates of change for $Y(\mathbf{s})$ and $X(\mathbf{s})$ at a location $\mathbf{s}$, replace $dEY(\mathbf{s})$ with the directional derivative for the response $D_{\mathbf{u}}Y(\mathbf{s})$ and replace $dX(\mathbf{s})$ with the directional derivative for the covariate $D_{\mathbf{u}}X(\mathbf{s})$. The resulting quantity, $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, is a well defined spatial process we refer to as the *local directional sensitivity process.* This local directional sensitivity is similar to β_1, informing about the response-covariate relationship, but will be unique to each location $\mathbf{s}$ and each direction $\mathbf{u}$.

Note, (1.4) defines a model for the response variable that can immediately be used to draw samples of the directional derivatives $D_{\mathbf{u}}Y(\mathbf{s})$ as described in Banerjee et al. (2003). However, it has not yet defined a model for the covariate $X(\mathbf{s})$, and therefore there is no distribution for $D_{\mathbf{u}}X(\mathbf{s})$ and no joint model for the pair of directional derivatives. The previous literature regarding spatial gradient processes considered only a univariate spatial process, so in order to examine the local directional sensitivity process it is necessary to extend the existing methodology to accommodate multivariate Gaussian processes.

In summary, the contribution of this thesis will be to extend the directional derivative methodology to multivariate Gaussian processes, enabing post model fitting sensitivity analysis. The procedure assumes that researchers have first fit a parent spatial model that they feel describes the data well. Standard inference procedures can then be complimented through post model fitting examination of the gradient processes. The posterior predictive distributions for the multivariate directional derivatives are completely determined by the parameters in the parent model,

requiring no additional model fitting to draw samples. Gradient-based processes, including the local directional sensitivity process, can then be computed and analyzed in order to glean additional insight into the spatial variable relationships. In particular, these processes assess the spatial variation in relationships without requiring a spatially varying coefficient model.

Chapter 2 extends the existing spatial gradient theory to accommodate bivariate Gaussian process models. The local directional sensitivity process and another gradient-derived process called the *spatial angular discrepancy process* are defined here and illustrated for a simple simulation example. The processes are shown to exhibit desirable properties and interpretable distributions, with the local directional sensitivity at each location centered at the global relationship parameter β_1.

Chapter 3 discusses multiple approaches to accomodating non-Gaussian responses and derives gradient theory under each model. To demonstrate post model fitting gradient analysis for a spatial GLM we turn to a plant species dataset coming from the Cape Floristic Region in South Africa. Presence-absence is modeled for a particular species given elevation, and the resulting gradient analysis illuminates behavior that would otherwise go undetected. To illustrate a non-Gaussian response that does not fit a GLM framework, the gradient analysis is then applied to output from a log-Gaussian Cox process used to jointly model point patterns of tree locations and elevation in Duke Forest. Specifically, the post model fitting analysis is used to compare spatial variation in two tree species' relationships with elevation.

Chapter 4 considers a few forms for the coregionalization model as a means to model multivariate Gaussian processes when multiple spatial responses and/or covariates are of interest. The local directional sensitivities are shown to exhibit different properties for each coregionalization specification, that may or may not be desirable depending on the application. Distribution theory is outlined for each of the three multivariate possibilities and is illustrated using a plant trait and climate

dataset from the Cape Floristic Region in South Africa. The dataset allows for analysis of a single response with multiple covariates as well as multiple responses with a single covariate.

Finally, Chapter 5 discusses relevant theory for several additional spatial modeling frameworks, including those outlined in subsection 1.1.2, further demonstrating the versatility of post model fitting gradient analyses.

2

Bivariate Spatial Gradients

2.1 Introduction

Spatial regression models commonly assume a linear relationship and make inference based on the coefficient assigned to the covariate. This coefficient describes the expected change in response, given a unit change in covariate, providing a global gradient. However, gradient behavior is expected to vary locally over the study region, and such local, or second-order, behavior can be studied through spatial gradient analysis.

Assuming a single response and single covariate of interest whose surfaces are spatially smooth, then both can be treated as realizations from a stochastic process. It may be believed that the rate of change of the covariate impacts the rate of change of the response and that the relationship between the two surfaces may differ across the domain. An example of this setting would consider species response to climate. For instance, ecologists are increasingly interested in characterizing how abundance and frequency of tree species relate to changes in climate (e.g. Canham and Thomas, 2010; Thuiller et al., 2004; Thomas, 2010). Investigating the response of

plants to climate can be informative about the expected effects of climate change on range distributions. For instance, we might want to learn how temperature gradients affect abundance gradients. Another example might consider how, across a region, phenological traits of a plant species, e.g., leaf size, leaf thickness, and leaf length-leaf diameter ratio, respond to moisture, soil type, and topography.

The contribution of this chapter is to extend the existing gradient theory to accommodate this sort of bivariate analysis by modeling the response and covariate jointly. If we work in a hierarchical modeling framework and fit this joint model, then corresponding gradients for the spatial surfaces can be sampled simultaneously from the joint predictive distribution. Under a significant regression relationship it is not sensible to investigate the gradient behavior of the surfaces marginally. Suitable comparison between the gradient surfaces illustrates how sensitive the response surface is to the covariate surface, as well as the strength of this relationship. The former is accomplished through comparison between the directions of the maximum gradient at a given location; the latter requires consideration of their directional derivatives relative to one another. In particular, we introduce two new spatial processes, a local directional sensitivity process and a spatial angular discrepancy process. These inferential tools are developed and carried out on simulated data in the context of a customary geostatistical model (Banerjee et al., 2004).

In section 2.2 the formal distribution theory for the spatial gradients is extended to the multivariate case. Section 2.3 outlines the modeling framework for the simulation example. This section also defines the two processes of interest, namely the local directional sensitivity process and the spatial angular discrepancy process. Section 2.4 provides a simulated example with a multivariate Gaussian process setup as a proof of concept. Finally, section 2.5 summarizes the contributions of the chapter.

2.2 Distribution Development

In this section we review the definitions and distributions presented in Banerjee et al. (2003) and extend these ideas to consider a multivariate Gaussian process. We assume locations $\mathbf{s} \in \mathbb{R}^2$, 2-dimensional Euclidean space, however extension to a generic d-dimensional setting is straightforward. The process is assumed, for convenience, to be (weakly) stationary such that the covariance function, $Cov(Y(\mathbf{s}), Y(\mathbf{s}'))$, depends only on the separation vector $\boldsymbol{\delta} = \mathbf{s} - \mathbf{s}'$. In fact, in our examples we adopt isotropic covariance functions that depend only on the length of the separation vector, $||\boldsymbol{\delta}||$.

Consider two surfaces $\{(Y(\mathbf{s}), X(\mathbf{s})) : \mathbf{s} \in \mathbb{R}^2\}$ drawn from a joint Gaussian process specified such that $X(\mathbf{s})$ has constant mean, say α_0, and covariance function $G(\boldsymbol{\delta})$. Given $X(\mathbf{s})$, $Y(\mathbf{s})$ has mean $\beta X(\mathbf{s})$ and covariance function $K(\boldsymbol{\delta})$. Observed at a set of locations $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))$, we write: $\mathbf{Y}|\mathbf{X} \sim N(\beta\mathbf{X}, K(\cdot))$ with $\mathbf{X} \sim N(\alpha_0, G(\cdot))$. Considered jointly, we have:

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_0\beta\mathbf{1} \\ \alpha_0\mathbf{1} \end{pmatrix}, \begin{pmatrix} K(\cdot) + \beta^2 G(\cdot) & \beta G(\cdot) \\ \beta G(\cdot) & G(\cdot) \end{pmatrix} \right)$$

where $G(\cdot)$ and $K(\cdot)$ are matrices of the covariance functions with entry i, j evaluated at $\boldsymbol{\delta} = \mathbf{s}_i - \mathbf{s}_j$.

We follow the notation and theory in Banerjee et al. (2003). Assume mean square differentiable processes $Y(\mathbf{s})$ and $X(\mathbf{s})$. That is, for $Y(\mathbf{s})$, at $\mathbf{s}_0$ there exists a vector $\nabla_Y(\mathbf{s}_0)$ such that for any scalar h and any unit vector $\mathbf{u}$, $Y(\mathbf{s}_0 + h\mathbf{u}) = Y(\mathbf{s}_0) + h\mathbf{u}^T\nabla_Y(\mathbf{s}_0) + r(\mathbf{s}_0, h\mathbf{u})$ where $r(\mathbf{s}_0, h\mathbf{u}) \to 0$ in the L_2 sense as $h \to 0$. Similarly, for $X(\mathbf{s})$.

In particular, define the finite difference processes at scale h in direction $\mathbf{u}$:

$$Y_{\mathbf{u},h}(\mathbf{s}) = \frac{Y(\mathbf{s} + h\mathbf{u}) - Y(\mathbf{s})}{h}$$

$$X_{\mathbf{u},h}(\mathbf{s}) = \frac{X(\mathbf{s} + h\mathbf{u}) - X(\mathbf{s})}{h}$$

where $\mathbf{u}$ is a unit vector. Taking the limit as h tends to 0, Banerjee et al. (2003) define the directional derivative processes in the direction $\mathbf{u}$:

$$D_{\mathbf{u}}Y(\mathbf{s}) = \lim_{h \to 0} Y_{\mathbf{u},h}(\mathbf{s}) = \mathbf{u}'\nabla_Y(\mathbf{s})$$

$$D_{\mathbf{u}}X(\mathbf{s}) = \lim_{h \to 0} X_{\mathbf{u},h}(\mathbf{s}) = \mathbf{u}'\nabla_X(\mathbf{s})$$

where $\nabla_X(\mathbf{s}) = (D_{\mathbf{e}_1}X(\mathbf{s}), D_{\mathbf{e}_2}X(\mathbf{s}))'$ is the vector of directional derivatives in the orthonormal basis directions $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ for $\mathbb{R}^2$. We can study the directional derivative processes for any $\mathbf{u}$ by working with the basis set $\nabla_Y(\mathbf{s})$ and $\nabla_X(\mathbf{s})$.

From Banerjee et al. (2003), we know that if $Y(\mathbf{s})$ and $X(\mathbf{s})$ are stationary Gaussian processes, then the resulting marginal distributions involving the directional derivatives will be stationary Gaussian processes as well. Note, isotropy in the $Y(\mathbf{s})$ process does not induce isotropy in the $D_{\mathbf{u}}Y(\mathbf{s})$ process; only stationarity will be inherited. Similar to the discussion in their paper, we know by linearity that $(Y(\mathbf{s}), X(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}))'$ will be a stationary multivariate Gaussian process. And then, by a standard limiting moment generating function argument, $(Y(\mathbf{s}), X(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}), D_{\mathbf{u}}X(\mathbf{s}))'$ will also be a stationary multivariate Gaussian process.

To explicitly provide the joint distribution, we derive the cross covariance structure by examining pair-wise covariances between the response and covariate processes and their directional derivatives. For notational convenience, write the marginal covariance function of $Y(\mathbf{s})$ to be $\tilde{K}(\cdot) = K(\cdot) + \beta^2 G(\cdot)$. Assume the $Y(\mathbf{s})$ and $X(\mathbf{s})$ processes are mean-zero, setting $\alpha_0 = 0$, since in practice the gradients are calculated for the mean-zero residual process. If $E(Y(\mathbf{s})) = 0$, then $E(D_{\mathbf{u}}Y(\mathbf{s})) = 0$, so the joint processes will all be mean-zero. We calculate the covariances associated with the directional derivatives by taking the limits of the covariances corresponding to the analogous finite difference process.

The covariances for the response surface are derived in Banerjee et al. (2003):

$$Cov(Y_{\mathbf{u},h}(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \frac{2\tilde{K}(\boldsymbol{\delta}) - \tilde{K}(\boldsymbol{\delta} + h\mathbf{u}) - \tilde{K}(\boldsymbol{\delta} - h\mathbf{u})}{h^2}$$

$$Cov(D_{\mathbf{u}}Y(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}')) = \lim_{h\to 0} Cov(Y_{\mathbf{u},h}(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = -\mathbf{u}'\Omega_{\tilde{K}}\mathbf{u}$$

$$Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \frac{\tilde{K}(\boldsymbol{\delta} - h\mathbf{u}) - \tilde{K}(\boldsymbol{\delta})}{h}$$

$$Cov(Y(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}')) = \lim_{h\to 0} Cov(Y(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = D_{\mathbf{u}}\tilde{K}(-\boldsymbol{\delta})$$

and covariances for the covariate surface are analogous:

$$Cov(X_{\mathbf{u},h}(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = \frac{2G(\boldsymbol{\delta}) - G(\boldsymbol{\delta} + h\mathbf{u}) - G(\boldsymbol{\delta} - h\mathbf{u})}{h^2}$$

$$Cov(D_{\mathbf{u}}X(\mathbf{s}), D_{\mathbf{u}}X(\mathbf{s}')) = \lim_{h\to 0} Cov(X_{\mathbf{u},h}(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = -\mathbf{u}'\Omega_{G}\mathbf{u}$$

$$Cov(X(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = \frac{G(\boldsymbol{\delta} - h\mathbf{u}) - G(\boldsymbol{\delta})}{h}$$

$$Cov(X(\mathbf{s}), D_{\mathbf{u}}X(\mathbf{s}')) = Cov(X(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = D_{\mathbf{u}}G(-\boldsymbol{\delta})$$

To fully describe the joint distribution we derive the covariances between response and covariate surfaces similarly:

$$Cov(Y(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = \frac{\beta G(\boldsymbol{\delta} - h\mathbf{u}) - \beta G(\boldsymbol{\delta})}{h}$$

$$Cov(Y(\mathbf{s}), D_{\mathbf{u}}X(\mathbf{s}')) = \lim_{h\to 0} Cov(Y(\mathbf{s}), X_{\mathbf{u},h}(\mathbf{s}')) = \beta D_{\mathbf{u}}G(-\boldsymbol{\delta})$$

$$Cov(X(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \frac{\beta G(\boldsymbol{\delta} - h\mathbf{u}) - \beta G(\boldsymbol{\delta})}{h}$$

$$Cov(X(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}')) = \lim_{h\to 0} Cov(X(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}')) = \beta D_{\mathbf{u}}G(-\boldsymbol{\delta})$$

$$Cov(X_{\mathbf{u},h}(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}') = \frac{2\beta G(\boldsymbol{\delta}) - \beta G(\boldsymbol{\delta} + h\mathbf{u}) - \beta G(\boldsymbol{\delta} - h\mathbf{u})}{h^2}$$

$$Cov(D_{\mathbf{u}}X(\mathbf{s}), D_{\mathbf{u}}Y(\mathbf{s}')) = \lim_{h\to 0} Cov(X_{\mathbf{u},h}(\mathbf{s}), Y_{\mathbf{u},h}(\mathbf{s}') = -\beta \mathbf{u}'\Omega_G(\boldsymbol{\delta})\mathbf{u}$$

where $(\Omega_G(\boldsymbol{\delta}))_{ij} = \partial^2 G(\boldsymbol{\delta})/\partial\boldsymbol{\delta}_i\partial\boldsymbol{\delta}_j$ and $D_{\mathbf{u}}G(\boldsymbol{\delta}) = \lim_{h\to 0}(G(\boldsymbol{\delta} - h\mathbf{u}) - G(\boldsymbol{\delta}))/h$.

Relationships between the response surface, the covariate surface and their corresponding directional derivative surfaces can be described through the 6-dimensional multivariate stationary Gaussian process $\mathbf{Z}(\mathbf{s}) = (Y(\mathbf{s}), X(\mathbf{s}), \nabla_Y(\mathbf{s}), \nabla_X(\mathbf{s}))'$. Using the covariances calculated above, the associated cross-covariance matrix for $\mathbf{Z}$ will be:

$$V_{\mathbf{Z}}(\boldsymbol{\delta}) = \begin{pmatrix} \tilde{K}(\boldsymbol{\delta}) & \beta G(\boldsymbol{\delta}) & -\nabla\tilde{K}(\boldsymbol{\delta})' & -\beta\nabla G(\boldsymbol{\delta})' \\ \beta G(\boldsymbol{\delta}) & G(\boldsymbol{\delta}) & -\beta\nabla G(\boldsymbol{\delta})' & -\nabla G(\boldsymbol{\delta})' \\ \nabla\tilde{K}(\boldsymbol{\delta}) & \beta\nabla G(\boldsymbol{\delta}) & -H_{\tilde{K}}(\boldsymbol{\delta}) & -\beta H_G(\boldsymbol{\delta}) \\ \beta\nabla G(\boldsymbol{\delta}) & \nabla G(\boldsymbol{\delta}) & -\beta H_G(\boldsymbol{\delta}) & -H_G(\boldsymbol{\delta}) \end{pmatrix}$$

$$= \begin{pmatrix} K(\boldsymbol{\delta}) + \beta^2 G(\boldsymbol{\delta}) & \beta G(\boldsymbol{\delta}) & -\nabla K(\boldsymbol{\delta})' - \beta^2\nabla G(\boldsymbol{\delta})' & -\beta\nabla G(\boldsymbol{\delta})' \\ \beta G(\boldsymbol{\delta}) & G(\boldsymbol{\delta}) & -\beta\nabla G(\boldsymbol{\delta})' & -\nabla G(\boldsymbol{\delta})' \\ \nabla K(\boldsymbol{\delta}) + \beta^2\nabla G(\boldsymbol{\delta}) & \beta\nabla G(\boldsymbol{\delta}) & -H_K(\boldsymbol{\delta}) - \beta^2 H_G(\boldsymbol{\delta}) & -\beta H_G(\boldsymbol{\delta}) \\ \beta\nabla G(\boldsymbol{\delta}) & \nabla G(\boldsymbol{\delta}) & -\beta H_G(\boldsymbol{\delta}) & -H_G(\boldsymbol{\delta}) \end{pmatrix}$$

where $\nabla K(\boldsymbol{\delta})$ is a 2×1 gradient vector associated with $K(\boldsymbol{\delta})$, and $H_K(\boldsymbol{\delta})$ is the 2×2 Hessian matrix associated with $K(\boldsymbol{\delta})$.

For $\boldsymbol{\delta} = \mathbf{0}$, we have a block diagonal local covariance matrix:

$$V_{\mathbf{Z}}(\mathbf{0}) = \begin{pmatrix} K(\mathbf{0}) + \beta^2 G(\mathbf{0}) & \beta G(\mathbf{0}) & \mathbf{0}' & \mathbf{0}' \\ \beta G(\mathbf{0}) & G(\mathbf{0}) & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & -H_K(\mathbf{0}) - \beta^2 H_G(\mathbf{0}) & -\beta H_G(\mathbf{0}) \\ \mathbf{0} & \mathbf{0} & -\beta H_G(\mathbf{0}) & -H_G(\mathbf{0}) \end{pmatrix}.$$

Thus, at a location $\mathbf{s}$, the directional derivative surfaces will be correlated with one another, but neither will be correlated with either of the data surfaces. Intuitively, this makes sense since we would not expect the level of the surface at a location to be correlated with the rate of change at that location. Of course, since $(X(\mathbf{s}), Y(\mathbf{s}))'$ is a bivariate Gaussian process, the correlation between the rates of changes is not surprising.

The Matérn covariance is adopted below. It depends on a smoothness parameter ν which directly controls the mean square differentiability of process realizations (Stein, 1999). This is convenient since, again, the $Y(\mathbf{s})$ and $X(\mathbf{s})$ processes must be mean square differentiable for their associated directional derivative processes to be well defined. If we let $K(\cdot)$ and $G(\cdot)$ be Matérn with $\nu > 1$ then they are once (but not twice) mean square differentiable, and, if $\nu = 3/2$, the covariance functions are of the closed form $\sigma^2(1 + \phi||\boldsymbol{\delta}||)\exp(-\phi||\boldsymbol{\delta}||)$. We denote the parameters of $K(\cdot)$ as σ_y^2 and ϕ_y, and the parameters of $G(\cdot)$ as σ_x^2 and ϕ_x.

Under the Matérn covariance the components of the cross-covariance matrix will be $\nabla K(\boldsymbol{\delta}) = -\sigma_y^2\phi_y^2 \exp(-\phi_y||\boldsymbol{\delta}||)\boldsymbol{\delta}$, $(H_K(\boldsymbol{\delta}))_{ii} = -\sigma_y^2\phi_y^2 \exp(-\phi_y||\boldsymbol{\delta}||)(1-\phi_y\delta_i^2/||\boldsymbol{\delta}||)$, $(H_K(\boldsymbol{\delta}))_{ij} = \sigma_y^2\phi_y^3 \exp(-\phi_y||\boldsymbol{\delta}||)\delta_i\delta_j/||\boldsymbol{\delta}||$, and similar for $G(\cdot)$. Then, we have for $\boldsymbol{\delta} = \mathbf{0}$:

$$V_{\mathbf{Z}}(\mathbf{0}) = \begin{pmatrix} \sigma_y^2 + \beta^2\sigma_x^2 & \beta\sigma_x^2 & \mathbf{0}' & \mathbf{0}' \\ \beta\sigma_x^2 & \sigma_x^2 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} & (\sigma_y^2\phi_y^2 + \beta^2\sigma_x^2\phi_x^2)I_2 & (\beta\sigma_x^2\phi_x^2)I_2 \\ \mathbf{0} & \mathbf{0} & (\beta\sigma_x^2\phi_x^2)I_2 & (\sigma_x^2\phi_x^2)I_2 \end{pmatrix} \tag{2.1}$$

where I_2 is the 2×2 identity matrix. As above, $D_{\mathbf{e}_1}Y(\mathbf{s})$ and $D_{\mathbf{e}_1}X(\mathbf{s})$ will be correlated with one another, and similarly $D_{\mathbf{e}_2}Y(\mathbf{s})$ and $D_{\mathbf{e}_2}X(\mathbf{s})$ will be correlated with

one another, but all other pairings of the directional derivatives will be uncorrelated.

2.3 Model Fitting and Inference

2.3.1 Sampling Method

Following the modeling of the previous section, let $K(\boldsymbol{\delta}) = \sigma_y^2 \rho_y(\boldsymbol{\delta})$ and $G(\boldsymbol{\delta}) = \sigma_x^2 \rho_x(\boldsymbol{\delta})$, where the ρ_x and ρ_y are valid two-dimensional correlation functions. We work with the Matérn class of covariance functions parameterized by ϕ and ν with $\nu > 1$.

Let $\boldsymbol{\theta} = (\alpha_0, \beta_0, \beta_1, \sigma_x^2, \sigma_y^2, \phi_x, \phi_y, \nu_x, \nu_y)$. For locations $\mathbf{s_1}, \ldots, \mathbf{s_n}$, the overall likelihood can be written in terms of the conditional likelihoods

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}) &\propto L(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{X}) L(\mathbf{X}|\boldsymbol{\theta}) \\ &\propto (\sigma_x^2 \sigma_y^2)^{-n/2} |R_x(\phi_x, \nu_x)|^{-1/2} |R_y(\phi_y, \nu_y)|^{-1/2} \\ &\times \exp\left\{ -\frac{1}{2\sigma_x^2} (\mathbf{X} - \alpha_0 \mathbf{1})' R_x^{-1}(\phi_x, \nu_x)(\mathbf{X} - \alpha_0 \mathbf{1}) \right\} \\ &\times \exp\left\{ -\frac{1}{2\sigma_y^2} (\mathbf{Y} - (\beta_0 \mathbf{1} + \beta_1 \mathbf{X}))' R_y^{-1}(\phi_y, \nu_y)(\mathbf{Y} - (\beta_0 \mathbf{1} + \beta_1 \mathbf{X})) \right\} \end{aligned}$$

where $\mathbf{Y} = (Y(\mathbf{s_1}), \ldots, Y(\mathbf{s_n}))'$, $(R_x(\phi_x, \nu_x))_{ij} = \rho_x(\mathbf{s}_i - \mathbf{s}_j; \phi_x, \nu_x)$ and $(R_y(\phi_y, \nu_y))_{ij} = \rho_y(\mathbf{s}_i - \mathbf{s}_j; \phi_y, \nu_y)$. The likelihood could be equivalently written in its joint form, but the conditional form is more conducive to interpreting and implementing the gradient analysis.

We see that we have a low dimensional parametric model, with $\boldsymbol{\theta}$ only 9 dimensional. We utilize fairly non-informative priors for its components. For example, vague normal priors on $(\alpha_0, \beta_0, \beta_1)$, vague inverse Gamma priors on (σ_x^2, σ_y^2), vague Gamma priors on (ϕ_x, ϕ_y), and $U(1, 2)$ priors on (ν_x, ν_y). The prior on ν follows the suggestion of Stein (1999) and others who observe that distinguishing $\nu = 2$ from $\nu > 2$ would be very difficult in practice. This model is straight forward to fit in its

conditional form, for example using the 'spBayes' package in R (Finley et al., 2007). Thus, assume we now have posterior samples $\boldsymbol{\theta}_l^*, l = 1, \ldots, L$, from $f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X})$.

Once we have posterior samples of the parameters, we draw samples of the gradient vectors using composition since the posterior predictive distribution $f(\nabla_Y, \nabla_X|\mathbf{Y}, \mathbf{X}) = \int f(\nabla_Y, \nabla_X|\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) d\boldsymbol{\theta}$. The cross-covariance matrix derived earlier allows us to immediately write the joint multivariate normal distribution given $\boldsymbol{\theta}$, which can be evaluated at each sample $\boldsymbol{\theta}_l^*$. Based on this joint distribution, standard multivariate normal theory allows us to write down the desired conditional distributions needed to draw from the predictive distribution.

For an unobserved location $\mathbf{s}_0$, obtaining draws of the gradient vectors is again done via the predictive distribution. The cross covariance matrix derived enables us to write the joint distribution, from which we can derive the conditional distribution $f(\nabla_Y(\mathbf{s}_0), \nabla_X(\mathbf{s}_0)|\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta})$. If interest is also in the values of the $Y(\mathbf{s})$ and $X(\mathbf{s})$ surfaces at the new location, we would derive the conditional distribution $f(Y(\mathbf{s}_0), X(\mathbf{s}_0), \nabla_Y(\mathbf{s}_0), \nabla_X(\mathbf{s}_0)|\mathbf{Y}, \mathbf{X}, \boldsymbol{\theta})$, which is again straight forward given the cross covariance matrix and allows joint prediction of the surfaces and their gradients at the new location.

If we want $Y(\mathbf{s})$ and $X(\mathbf{s})$ to be adjusted based on some fixed covariates, then we simply introduce such covariates into the mean functions of the model. We create a spatial random effects model:

$$Y(\mathbf{s})|X(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + \mathbf{T}_y(\mathbf{s})'\boldsymbol{\gamma}_y + w_y(\mathbf{s}) + \epsilon(\mathbf{s})$$
$$X(\mathbf{s}) = \alpha_0 + \mathbf{T}_x(\mathbf{s})'\boldsymbol{\gamma}_x + w_x(\mathbf{s})$$

where $\mathbf{T}_y(\mathbf{s})$ and $\mathbf{T}_x(\mathbf{s})$ are vectors of covariates used to explain the $Y(\mathbf{s})$ and $X(\mathbf{s})$ surfaces respectively, with coefficients $\boldsymbol{\gamma}_y$ and $\boldsymbol{\gamma}_x$; $w_y(\mathbf{s})$ and $w_x(\mathbf{s})$ are independent mean-zero stationary Gaussian processes with parameters $\sigma_x^2, \sigma_y^2, \phi_x, \phi_y, \nu_x, \nu_y$ as before; and $\epsilon(\mathbf{s})$ is a Gaussian white-noise process with variance τ^2 intended to capture

measurement error or microscale variability in the response. The $X(\mathbf{s})$ process is assumed to be a fully spatial model (no nugget effect), such as might be used for elevation, temperature, or pollutant level.

With $\boldsymbol{\theta}$ now extended to $\boldsymbol{\theta} = (\alpha_0, \beta_0, \beta_1, \boldsymbol{\gamma}_x, \boldsymbol{\gamma}_y, \sigma_x^2, \sigma_y^2, \phi_x, \phi_y, \nu_x, \nu_y)$, the likelihood for locations $\mathbf{s_1}, \ldots, \mathbf{s_n}$ above can be trivially revised. Prior selection for the parameters will be similar to the previous example. Again, this model can be implemented using `'spBayes'`, and draws of the gradients will rely on the posterior predictive distribution, which can be calculated as before using the derived cross-covariance matrix.

2.3.2 Local Directional Sensitivity Process

At a given location $\mathbf{s}$ there may additionally be interest in the ratio of directional derivatives, $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, corresponding to the relative rates of change in the two surfaces in direction $\mathbf{u}$. This quantity is analogous to dy/dx in more standard calculus applications as well as sensitivity functions studied in sensitivity analysis (Tomovic and Vukobratovic, 1972). With this in mind, we refer to the resulting spatial process as the *local directional sensitivity process*. The choice of $\mathbf{u}$ will depend on the application being considered. For example, this direction may correspond to latitudinal direction, an elevation direction, or to an environmental feature expected to impact the response. Large values of this process would suggest that the change in covariate surface has a high impact on the response surface.

The joint process defined in Section 2.2 can be equivalently represented as a

spatial random effects model (excluding nugget effects):

$$Y(\mathbf{s})|X(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s})$$
$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$
$$w_y(\mathbf{s}) \sim GP(0, K(\boldsymbol{\delta}))$$
$$w_x(\mathbf{s}) \sim GP(0, G(\boldsymbol{\delta}))$$

where $w_y(\mathbf{s})$ and $w_x(\mathbf{s})$ are independent processes. With this notation we can write the unconditional response surface and corresponding directional derivative process as follows:

$$Y(\mathbf{s}) = \beta_0 + \beta_1\alpha_0 + \beta_1 w_x(\mathbf{s}) + w_y(\mathbf{s})$$
$$D_{\mathbf{u}}Y(\mathbf{s}) = \beta_1 D_{\mathbf{u}} w_x(\mathbf{s}) + D_{\mathbf{u}} w_y(\mathbf{s})$$

The local directional sensitivity process can then be written

$$\frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \frac{\beta_1 D_{\mathbf{u}} w_x(\mathbf{s}) + D_{\mathbf{u}} w_y(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} = \beta_1 + \frac{D_{\mathbf{u}} w_y(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \tag{2.2}$$

We see that the multiplicative parameter, β_1, defining the overall relationship between the $X(\mathbf{s})$ and $Y(\mathbf{s})$ processes serves to center the local directional sensitivity process.

As mentioned in Majumdar et al. 2006, if we consider $Y(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + \epsilon(\mathbf{s})$, then one could write $\beta_1 = dE(Y(\mathbf{s}))/dX(\mathbf{s})$; i.e., β_1 is describing the rate of change in $E(Y(\mathbf{s}))$ relative to changes in $X(\mathbf{s})$. Again, at any location the local directional sensitivity will be centered at the global (non-directional) derivative ratio $dE(Y(\mathbf{s}))/dX(\mathbf{s})$ plus some (directional) spatial noise. In this way, the local directional sensitivity process is describing the spatial variation in the relative rates of change between $X(\mathbf{s})$ and $Y(\mathbf{s})$. This is analogous to modeling adopting spatially varying coefficients, $\beta(\mathbf{s})$, (Gelfand et al., 2003) but is arguably a simpler context since the derivatives require no additional model fitting. In addition, consideration

of directional perspectives using this gradient approach allows for inference distinct from what one can learn from a non-directional $\beta(\mathbf{s})$ parameter.

By noting that $D_{\mathbf{u}}w_y(\mathbf{s})/D_{\mathbf{u}}w_x(\mathbf{s})$ is a ratio of independent mean zero normal random variables, at each location the directional spatial noise is a Cauchy random variable with scale equal to the ratio of the respective standard deviations, $SD(D_{\mathbf{u}}w_y(\mathbf{s}))/SD(D_{\mathbf{u}}w_x(\mathbf{s}))$. If $\mathbf{u} = (u_1, u_2)$, then

$$SD(D_{\mathbf{u}}w_y(\mathbf{s})) = \sqrt{u_1^2(-H_K(0)_{1,1}) + u_2^2(-H_K(0)_{2,2})}$$

If $Y(\mathbf{s})$ is isotropic, then $H_K(0) = c_o I_2$ (Banerjee et al., 2003) and $SD(D_{\mathbf{u}}w_y(\mathbf{s})) = \sqrt{u_1^2 + u_2^2}\sqrt{-c_0} = \sqrt{-c_0}$. If $K(\cdot)$ and $G(\cdot)$ are Matérn with $\nu = 3/2$, then $SD(D_{\mathbf{u}}w_i(\mathbf{s})) = \sigma_i\phi_i$, and the scale for the Cauchy distribution will be $\sigma_y\phi_y/\sigma_x\phi_x$. When the respective standard deviations are equal, the scale will be 1 and the directional derivative ratio will have a standard Cauchy distribution.

In fact, the collection of directional derivative ratios form a well defined spatial stochastic process which would naturally be called a spatial Cauchy process (see Appendix A.1 for details). For any set of locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ the joint distribution is well defined. For example, consider two locations $\mathbf{s}$ and $\mathbf{s}'$, simplifying the notation for clarity:

$$\begin{aligned} P(\frac{D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})} < r_1, \frac{D_{\mathbf{u}}w_y(\mathbf{s}')}{D_{\mathbf{u}}w_x(\mathbf{s}')} < r_2) &= P(\frac{n_1}{m_1} < r_1,\ \frac{n_2}{m_2} < r_2) \\ &= P(n_1 < r_1 m_1,\ n_2 < r_2 m_2,\ m_1 > 0,\ m_2 > 0) \\ &+ P(n_1 < r_1 m_1,\ n_2 > r_2 m_2,\ m_1 > 0,\ m_2 < 0) \\ &+ P(n_1 > r_1 m_1,\ n_2 < r_2 m_2,\ m_1 < 0,\ m_2 > 0) \\ &+ P(n_1 > r_1 m_1,\ n_2 > r_2 m_2,\ m_1 < 0,\ m_2 < 0) \end{aligned}$$

In turn, each of these terms can be computed as an integral involving normal densi-

ties. For example, the first term can be written as follows:

$$P(n_1 < r_1 m_1, n_2 < r_2 m_2, m_1 > 0, m_2 > 0) =$$

$$\int_0^\infty \int_0^\infty \int_{-\infty}^{r_1 m_1} \int_{-\infty}^{r_2 m_2} f_K(n_1,\ n_2) f_G(m_1,\ m_2) dn_2 dn_1 dm_2 dm_1 \tag{2.3}$$

where $f_K(n_1, n_2) = f_K(D_u w_y(\mathbf{s}), D_u w_y(\mathbf{s}'))$ is the bivariate normal density for the $D_u w_y(\mathbf{s})$ Gaussian process with parent covariance function $K(\cdot)$ evaluated at $\mathbf{s}$ and $\mathbf{s}'$, and similar for $f_G(m_1, m_2)$. Recall that a univariate Cauchy distribution can be defined as a scale mixture of normals (Andrews and Mallows, 1974). If we write the corresponding cdf as $F(r) = \int_0^\infty \int_{-\infty}^{rm} \phi(n)\phi(m) dn dm$, then the form in (2.3) is evidently similar and can be regarded as a bivariate analogue. In this way, the spatial Cauchy process is defined such that at any location the distribution is a univariate Cauchy distribution and for any set of locations the distribution is a sum of integrals of a similar form.

In the data analysis examples we consider a fixed direction $\mathbf{u}$ and draw samples of the directional derivative ratio at each location across a region. Although we cannot show mean square continuity for this surface, by imposing additional smoothness conditions on the covariance functions for $w_y(\mathbf{s})$ and $w_x(\mathbf{s})$, we can argue that this surface will be almost surely continuous using results from Kent (1989), following the development in Banerjee and Gelfand (2003). We omit the details.

2.3.3 Spatial Angular Discrepancy Process

At any given location there may be interest not only in the magnitude of the gradients in various directions, but also in the direction at which the maximum gradients are achieved. Stronger alignment between the directions of maximum gradients would suggest a stronger relationship between the response and covariate surfaces.

Consider the covariate process $X(\mathbf{s})$. At location $\mathbf{s}$ the maximum gradient will be achieved in the direction described by the unit vector $\mathbf{u}_X^* = \nabla_X(\mathbf{s})/||\nabla_X(\mathbf{s})||$, and

the directional derivative in the direction of maximum gradient will be $D_{\mathbf{u}_X^*}X(\mathbf{s}) = ||\nabla_X(\mathbf{s})||$ (Banerjee et al., 2003). We can similarly consider the direction of maximum gradient for the response surface which will occur for $\mathbf{u}_Y^* = \nabla_Y(\mathbf{s})/||\nabla_Y(\mathbf{s})||$. In applications, researchers may be interested in the behavior of the response surface, or its relative rate of change, in the direction $\mathbf{u}_X^*$. That is, there may be interest in $D_{\mathbf{u}_X^*}Y(\mathbf{s}) = \nabla_X(\mathbf{s})'\nabla_Y(\mathbf{s})/||\nabla_X(\mathbf{s})||$ as well as $D_{\mathbf{u}_X^*}Y(\mathbf{s})/D_{\mathbf{u}_X^*}X(\mathbf{s}) = (\nabla_X(\mathbf{s})'\nabla_Y(\mathbf{s}))/(\nabla_X(\mathbf{s})'\nabla_X(\mathbf{s}))$.

At a location $\mathbf{s}$ the magnitude of the maximum gradient, $||\nabla_X(\mathbf{s})||$, will be the square root of a sum of squared independent normal random variables; as such, this quantity will have a Chi distribution with $d = 2$ degrees of freedom, possibly scaled by some factor. If the process is isotropic, then $||\nabla_X(\mathbf{s})||$ will have a Chi distribution with $d = 2$ degrees of freedom scaled by $SD(D_{\mathbf{e}_1}X(\mathbf{s})) = SD(D_{\mathbf{e}_2}X(\mathbf{s}))$. For the Matérn covariance ($\nu = 3/2$) structure this scaling factor will be equal to $\sigma_x\phi_x$.

The unit vector describing the direction of max gradient for the covariate surface can equivalently be described by an angle $\theta_X(\mathbf{s})$ such that $\tan(\theta_X(\mathbf{s})) = D_{(0,1)}X(\mathbf{s})/D_{(1,0)}X(\mathbf{s})$, and similarly for the response surface. The angle θ can take values from $-\pi$ to π, so inversion of tan must be done with care. Since tan has a period of only π, the inverse function is typically taken to be arctan*. (arctan$^*(S/C)$ is $\arctan(S/C)$ if $C > 0$, $S \geq 0$; $\pi/2$ if $C = 0$, $S > 0$; $\arctan(S/C) + \pi$ if $C < 0$; $\arctan(S/C) + 2\pi$ if $C \geq 0$, $S < 0$; and undefined if $C = 0$, $S = 0$ (Jammalamadaka and Sengupta, 2001).)

As smooth functions of a well defined spatial process, the directions of maximum gradient $(\theta_X(\mathbf{s}), \theta_Y(\mathbf{s}))$ define a bivariate projected Gaussian process, analogous to the univariate projected Gaussian processes described in Wang (2013). Marginally $\theta_X(\mathbf{s})$ and $\theta_Y(\mathbf{s})$ will each be a projected Gaussian process (Wang, 2013). Assuming the Matérn ($\nu = 3/2$) covariance structure we can derive the joint distribution of the

two angles at a given location $\mathbf{s}$:

$$f(\theta_X(\mathbf{s}), \theta_Y(\mathbf{s})) = \begin{cases} C\left(B + \frac{\sqrt{2\pi}\tilde{A}(\mathbf{s})}{(ac-\tilde{A}^2(\mathbf{s}))^{3/2}} L(0,0,\sqrt{\frac{\tilde{A}^2(\mathbf{s})}{ac}}) + \frac{\phi(0)}{ac}\right) & , \tilde{A}(\mathbf{s}) > 0 \\ C\left(B + \frac{\sqrt{2\pi}\tilde{A}(\mathbf{s})}{(ac-\tilde{A}^2(\mathbf{s}))^{3/2}} \left(0.5 - L(0,0,\sqrt{\frac{\tilde{A}^2(\mathbf{s})}{ac}})\right) + \frac{\phi(0)}{ac}\right) & , \tilde{A}(\mathbf{s}) < 0 \end{cases}$$

where $C = \frac{1}{a(2\pi)^{3/2}\sqrt{|\Sigma|}}$, $B = \frac{\tilde{A}^2(\mathbf{s})\phi(0)}{ac(ac-\tilde{A}^2(\mathbf{s}))}$, $|\Sigma| = (\sigma_x^2\phi_x^2)^2(\sigma_y^2\phi_y^2)^2$, $a = 1/(\sigma_y^2\phi_y^2)$, $c = (\sigma_y^2\phi_y^2 + \beta^2\phi_x^2\sigma_x^2)/(\sigma_x^2\phi_x^2)$, $\tilde{A}(\mathbf{s}) = \sqrt{a}\beta\cos(\theta_X(\mathbf{s}) - \theta_Y(\mathbf{s}))$, and $L(0,0,\rho)$ is the zero mean bivariate normal cdf with correlation ρ and standard deviations equal to 1 evaluated at $(0,0)$. A brief derivation is provided in Appendix A.2. In addition, it is straightforward to show that at each location $\mathbf{s}$ the angles, $\theta_X(\mathbf{s})$ and $\theta_Y(\mathbf{s})$, will marginally be uniform on $(-\pi, \pi)$.

The above bivariate density is plotted in Figure 2.1 for $\beta = (\pm 0.05, \pm 0.5, \pm 1)'$, holding all other parameters constant: $\sigma_y = 1 = \sigma_x$ and $\phi_y = 1.05 = \phi_x$. When $\beta > 0$ the mass is concentrated around (θ_X, θ_Y) pairs that are equal; when $\beta < 0$ the mass is concentrated around pairs where $\theta_X = \theta_Y - \pi$. When $\beta = \pm 0.05$ the relationship between $X(\mathbf{s})$ and $Y(\mathbf{s})$ is weak, and the density is roughly uniform over all angle pairs. As the magnitude of β increases, the mass becomes increasingly concentrated around these respective values.

To compare the directions of maximum gradient calculated from the posterior distribution, we compute a "discrepancy" for $\theta_X(\mathbf{s})$ and $\theta_Y(\mathbf{s})$. Define: $disc(\mathbf{s}) = 1 - \cos(\theta_X(\mathbf{s}) - \theta_Y(\mathbf{s}))$. As a smooth function of a spatial process, this discrepancy is a well defined spatial process which we refer to as the *spatial angular discrepancy process*. When the maximum gradients occur in identical directions this process will have a value of zero, when they occur in opposite directions the process will have a value of two. In our analyses we consider the process across a region, plotting the posterior median surface.

In some areas the $X(\mathbf{s})$ or $Y(\mathbf{s})$ surface may be quite flat and there will be no

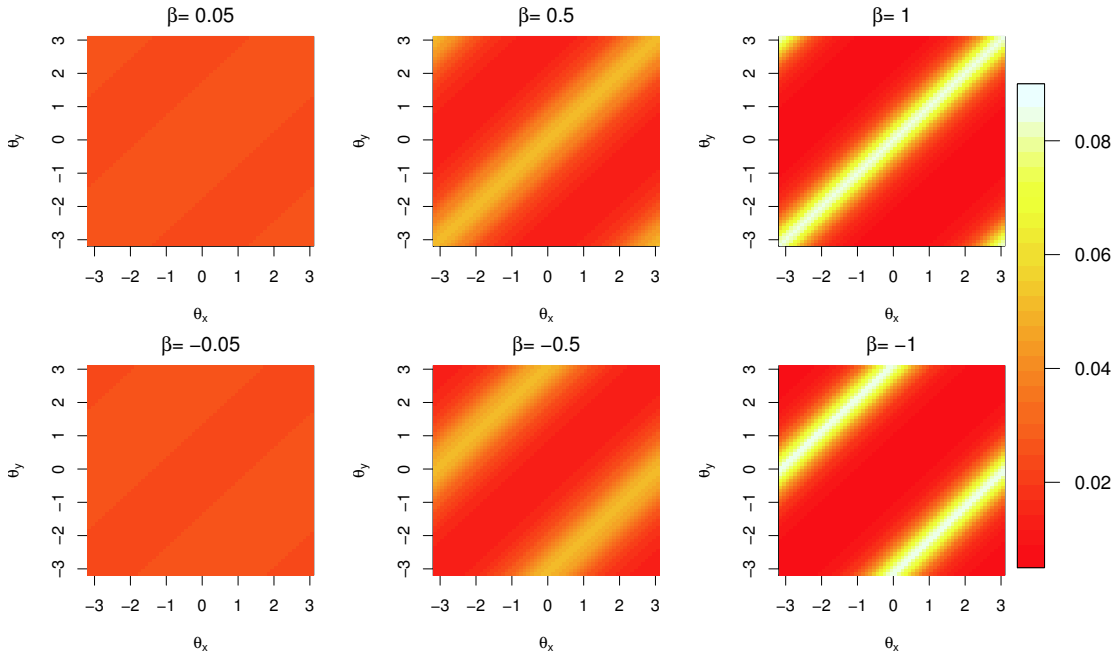

FIGURE 2.1: Bivariate density for $(\theta_X(\mathbf{s}), \theta_Y(\mathbf{s}))$ at a fixed $\mathbf{s}$ for varying values of β. All other parameters are set to the values used for simulation in Section 2.4.

direction with a gradient magnitude substantially larger than in the other directions. In these areas a large angular discrepancy may not be as meaningful as it would be in areas with larger gradient magnitudes. For this reason it is useful to examine these plots in tandem with plots of the local directional sensitivity process described in Section 2.3.2.

2.4 Simulation Example

We consider a simulation example to explore the behavior of gradient quantities in a controlled setting. From the foregoing development, in the context of spatial gradient analysis the quantities of interest will be directional derivatives. These derivatives are unobservable even in a simulation study with known parameters. Thus, assessing inference performance with regard to these quantities requires some novelty.

Recall that the gradient processes describe the shape and behavior of spatial sur-

faces. With a simulated dataset, we are able to draw a realization of the Gaussian process over a larger number of locations, allowing for fairly detailed understanding of the spatial surfaces. To avoid any unfair advantage that might come from the increased sample size, we use only a subset of the locations to fit the model and reserve the full set of locations for assessing the quality of our inference. Contour lines highlighting the shape of the surface are interpolated using the full set of locations. Conclusions made using gradients are then compared to those suggested by the contour lines to examine performance.

2.4.1 Data and Model

We simulate $X(\mathbf{s})$ a realization from a mean zero Gaussian process on $[0, 10] \times [0, 10]$, and $Y(\mathbf{s})|X(\mathbf{s})$ a realization from a Gaussian process with mean $\beta X(\mathbf{s})$. We use Matérn covariance functions setting $\nu = 3/2$ in order to capitalize on the resultant closed form. We simulate the Gaussian processes assuming Matérn covariance functions with parameters $\phi_x = 1.05 = \phi_y$, $\sigma_x^2 = 1 = \sigma_y^2$, $\beta_0 = 0 = \alpha_0$, and $\beta_1 = 0.5$. We draw a larger realization at 2000 locations, to allow for finer knowledge of the underlying surface, from which we consider a subset of 200 locations to be our "observations".

Treating $(Y(\mathbf{s}), X(\mathbf{s}))'$ as a multivariate Gaussian process, we fit a coregionalization model using the conditional parameterization, as described in Section 2.2. The model is fitted using the `'spbayes'` package in `R` by first fitting parameters for $X(\mathbf{s})$, then fitting parameters for $Y(\mathbf{s})|X(\mathbf{s})$. We obtain 2000 samples after a burn-in of 500 iterations and a thinning of every fifth iterate. We assume $\nu = 3/2$. Priors for the remaining parameters are: $\alpha_0 \sim N(0, 100), \beta_0 \sim N(0, 100), \beta_1 \sim N(0, 100), \phi_x, \phi_y \sim U(0.5, 10), \sigma_x^2, \sigma_y^2 \sim IG(2, 0.1)$ where $\alpha_0 = E(X(\mathbf{s}))$ and $\beta_0 + \beta_1 X(\mathbf{s}) = E(Y(\mathbf{s})|X(\mathbf{s}))$. Summaries of the posterior parameter samples are provided in Table 2.4.1.

We consider a region centered at the location $\mathbf{s}^* = (7.5, 6.5)$. The $X(\mathbf{s})$ and $Y(\mathbf{s})$

Table 2.1: Parameter estimates for the $[Y(\mathbf{s})|X(\mathbf{s})][X(\mathbf{s})]$ model.

Parameter	0.025	Mean	0.975	Truth
α_0	-0.7126	-0.0614	0.6214	0
β_0	0.2726	0.8179	1.3867	0
β_1	0.4685	0.5943	0.7202	0.5
σ_x^2	0.6725	1.0692	1.7390	1
ϕ_x	0.8242	1.0230	1.2124	1.05
σ_y^2	0.4911	0.8057	1.3341	1
ϕ_y	0.8572	1.0718	1.3116	1.05

values at this location are provided in Figure 2.2. (The full processes were realized on $[0, 10] \times [0, 10]$, but we only show a subregion here.) The interpolated surface and contour lines are produced using the full 2000 locations, while the circles indicate the subset of 200 locations used to predict the gradient.

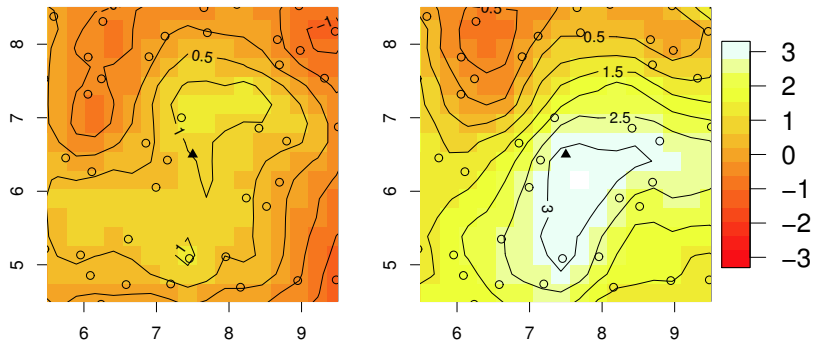

FIGURE 2.2: $X(\mathbf{s})$ (left) and $Y(\mathbf{s})$ (right) subregions around $\mathbf{s}^*$, where we estimate the gradient.

2.4.2 Local Directional Sensitivity Process

We are interested in the behavior of $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$. We consider $\mathbf{u} = (1, 0)$ and $\mathbf{u} = (0, 1)$. Since $D_{-\mathbf{u}}Y(\mathbf{s}) = -D_{\mathbf{u}}Y(\mathbf{s})$ (Banerjee et al., 2003), any discussion of the behavior in $\mathbf{u}$ direction implies the opposite behavior is occurring in the opposite direction. When applied to the ratios, this means that the local directional sensitivity process will be equal for $\mathbf{u}$ and $-\mathbf{u}$.

Returning to the region in Figure 2.2, to visualize the local directional sensitivity process we draw samples at a grid of 125 locations denoted as $\{\mathbf{s}_1^*, \ldots, \mathbf{s}_n^*\}$. We draw 2000 samples of $(\nabla_Y(\mathbf{s}_1^*), \ldots, \nabla_Y(\mathbf{s}_n^*), \nabla_X(\mathbf{s}_1^*), \ldots, \nabla_X(\mathbf{s}_n^*))'$ from the joint predictive distribution, again, given the fitting data at 200 observed locations. For each of these samples we calculate $D_{\mathbf{u}}Y(\mathbf{s}^*)/D_{\mathbf{u}}X(\mathbf{s}^*)$. We summarize the central behavior of these Cauchy quantities at a given location using the median value of the ratios.

For the two directions being considered, we plot the median predictive surface in Figure 2.3. Interpretation of these surfaces requires examination of the sign of the ratio as well as the magnitude. Magnitudes less than 1 suggest that the $X(\mathbf{s})$ surface is changing more rapidly than the $Y(\mathbf{s})$ surface; magnitudes greater than 1 suggest that the $Y(\mathbf{s})$ surface is changing more rapidly than the $X(\mathbf{s})$ surface; negative values suggest that one surface is increasing while the other decreases; positive values suggest that both surfaces are either increasing or decreasing.

The direction $\mathbf{u} = (1, 0)$ points towards the east. The corresponding ratio surface is provided in the left hand plot of Figure 2.3. There is a peak in the ratio surface around (6.75,6.25), suggesting that both surfaces are either decreasing or increasing and that the $Y(\mathbf{s})$ surface is doing so more rapidly. Referring back to Figure 2.2, the contour lines indicate that both surfaces are increasing at that location looking east, and that the $Y(\mathbf{s})$ surface is doing so more rapidly. The direction $\mathbf{u} = (0, 1)$ points towards the north. The corresponding ratio surface is provided in the right hand plot of Figure 2.3. There is a peak in the ratio surface around (8.75, 6), suggesting that both surfaces are either decreasing or increasing and that the $Y(\mathbf{s})$ surface is doing so more rapidly. Referring back to Figure 2.2, the contour lines indicate that both surfaces are increasing at that location looking east, and that the $Y(\mathbf{s})$ surface is doing so more rapidly.

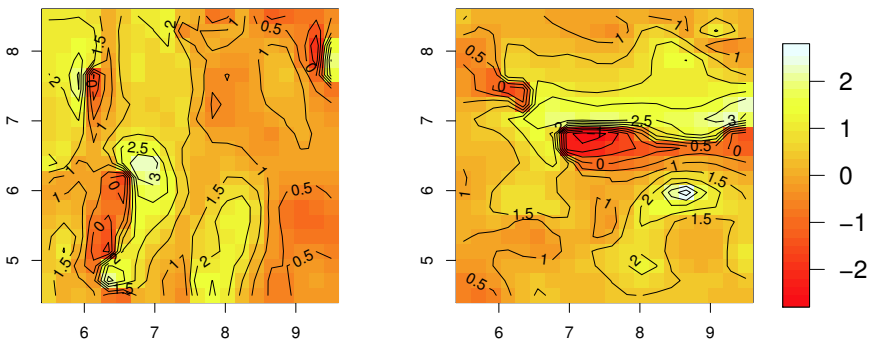

FIGURE 2.3: Posterior median of $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ in the directions $\mathbf{u} = (1, 0)$ (left) and $\mathbf{u} = (0, 1)$ (right).

2.4.3 Spatial Angular Discrepancy Process

Consider again the region in Figure 2.2 and the grid of 125 locations denoted as $\{\mathbf{s}_1^*, \ldots, \mathbf{s}_n^*\}$. Sampling gradients from the joint predictive distribution for $(\nabla_X(\mathbf{s}_i^*), \nabla_Y(\mathbf{s}_i^*))'$, we calculate the direction of maximum gradient as $\nabla_X(\mathbf{s}_i^*)/||\nabla_X(\mathbf{s}_i^*)||$ and $\nabla_Y(\mathbf{s}_i^*)/||\nabla_Y(\mathbf{s}_i^*)||$ for each sample gradient at each location in the figure. Denote these angles (in radians) as $\theta_X(\mathbf{s}_i^*)$ and $\theta_Y(\mathbf{s}_i^*)$ respectively.

We compute the discrepancy between these angles at each location as in Section 2.3.3 and provide the posterior median values in Figure 2.4. Most of the region has an associated discrepancy of 0, suggesting that both $X(\mathbf{s})$ and $Y(\mathbf{s})$ are typically increasing most rapidly in the same direction. However, there are a few small regions where the discrepancy peaks towards a value of 2, locations where the $X(\mathbf{s})$ and $Y(\mathbf{s})$ surfaces are increasing in nearly opposite directions.

2.5 Summary

We have extended previous theory on gradient analysis to allow for consideration of a bivariate process where one variable is treated as a response to the other variable. Consideration of the associated directional derivatives can be done jointly and results

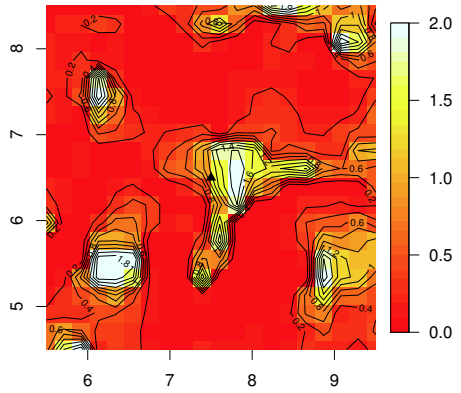

FIGURE 2.4: Posterior median $disc(\mathbf{s})$.

in a multivariate Gaussian process directly derivable from the model for the parent process. The models for the parent process are fit in a Bayesian framework and utilizing the posterior draws of the process parameters, all gradient analysis occurs post model fitting.

Using the directional derivatives, we proposed two derived processes that are useful for learning about the relationship between the response and covariate processes. The first is the local directional sensitivity process. This process captures local variation in the relationship between the two variables and provides deeper insight into their relative behavior. The second is the spatial angular discrepancy process, capturing the discrepancy between the directions in which the process surfaces are most rapidly increasing. Spatial plots of this discrepancy surface highlight regions of the domain where the two processes behave most similarly and most differently.

The current theory provides opportunity for several extensions. Many ecological data sets are observed at multiple time points, in part to see if the relationships between the variables of interest are changing over time. With this in mind, future work on gradient analyses may involve the incorporation of temporal effects. Ecological data sets also tend to have multiple responses and multiple covariates, any or all of which may have relationships that could be better highlighted through a gradient

analysis under a joint model. Finally, for many applications it will be important to consider novel non-Gaussian responses modeled through latent Gaussian process models.

3

Gradient Analysis for Non-Gaussian Responses

3.1 Introduction

While the methodology outlined in the previous chapter can readily accommodate spatial data with a continuous response and covariate, it is common to have data that naturally requires a non-Gaussian first stage for the model. Examples include binary or zero-inflated data, but there are countless others. Appropriate spatial models for these non-Gaussian responses include the class of spatial generalized linear models (GLMs), as well as others that similarly rely on a transformation of a latent Gaussian process. Computation of the directional derivative processes associated with these non-Gaussian models is made possible through the *spatial gradient chain rule*, defined in this chapter.

As a more basic spatial GLM example we use a dataset coming from a subregion of the Cape Floristic Region (CFR) in South Africa to present a spatial GLM gradient analysis with a binomial first stage. The analysis focuses on the spatial presence-absence behavior of a Protea species in relation to elevation. The relationship is explored at a regional level through examination of the spatial GLM parameters and

is explored at a more local level using the local directional sensitivity process and spatial angular discrepancy process.

Next, we consider a more novel non-Gaussian response using point pattern data from the Duke Forest. Conditional on elevation, point patterns for two tree species are each modeled as a log-Gaussian Cox process, with elevation modeled as a spatially varying intercept. The local directional sensitivity process and spatial angular discrepancy process provide inference regarding the relationship between point pattern intensity and elevation. One species illustrates a gradient analysis in the context of a non-significant relationship with elevation, and the other illustrates a gradient analysis in the context of a significant relationship with elevation.

The format of this chapter is as follows. The spatial gradient chain rule is outlined in Section 3.2. In Section 3.3 we discuss procedures for implementing the gradient analysis within a few spatial GLM frameworks. In Section 3.4 we detail the CFR data used for analysis, outline the binomial model and derive associated gradient theory. Results of the model fitting and gradient analyses are discussed with interpretations in the context of the binomial model. Section 3.5 provides the analysis of point pattern data from Duke Forest, illustrating gradient analysis for a log-Gaussian Cox process and comparing gradient analyses across significant and non-significant spatial regressions. Finally, Section 3.6 summarizes the results and contributions of the chapter.

3.2 Extensions to Non-Gaussian Data Models Using the Chain Rule

For data where the response variable is inherently non-Gaussian, spatial dependence is often captured through the use of a latent Gaussian surface. While the gradient analysis associated with these models will correspond to the latent Gaussian process, and not to the response of interest, differentiable transformations will allow for inference on the non-Gaussian response surface. Recall the *spatial gradi-*

ent chain rule presented in Majumdar et al. (2006): for $h(\cdot)$ differentiable on $\mathbb{R}^1$ and $W(\mathbf{s}) = h(V(\mathbf{s}))$, the directional derivative $D_{\mathbf{u}}W(\mathbf{s})$ exists and is given by $D_{\mathbf{u}}W(\mathbf{s}) = D_{\mathbf{u}}h(V(\mathbf{s})) = h'(V(\mathbf{s}))D_{\mathbf{u}}V(\mathbf{s})$. Here $V(\mathbf{s})$ is the latent Gaussian surface and the function $h(\cdot)$ describes its relationship to the non-Gaussian response.

For example, in the case of binary data $h^{-1}(\cdot)$ may be the link function relating the presence probability surface to a latent Gaussian process. Although the probability surface is non-Gaussian, the spatial gradient chain rule allows for consideration of its directional derivatives by examining the derivative of the inverse link function and the directional derivative for the latent Gaussian processes. These ideas are further detailed in the following examples.

3.3 Spatial Generalized Linear Models

A GLM replaces a model's typical Gaussian first stage with another member of the exponential family. The responses $Y(\mathbf{s})$ are assumed to be conditionally independent given the parameters, with distribution

$$\begin{aligned} f(Y(\mathbf{s})|\beta,\gamma) &= h(Y(\mathbf{s}),\gamma)\exp(\gamma(Y(\mathbf{s})\eta(\mathbf{s}) - \psi(\eta(\mathbf{s})))) \\ g(\eta(\mathbf{s})) &= X(\mathbf{s})\beta \end{aligned} \tag{3.1}$$

where g is a link function and γ is a dispersion parameter. These GLMs can be modified to allow for spatial dependence through the introduction of a zero-mean latent Gaussian process, commonly in the second stage of the hierarchical model. E.g., the expression in (3.1) above can be modified such that $g(\eta(\mathbf{s})) = X(\mathbf{s})\beta + w_z(\mathbf{s})$ where the spatial random effects $w_z(\mathbf{s})$ are realizations from a zero-mean Gaussian process. The latent process induces marginal spatial dependence in the response by encouraging the means of spatial variables at proximate locations to be similar (Banerjee et al., 2004).

The use of latent Gaussian processes when building spatial GLMs allows for

straight forward application of the developed spatial process gradient methodology for post model fitting analysis. Provided that the inverse link function $g^{-1}(\cdot)$ is differentiable, we can apply the spatial gradient chain rule to learn about the directional derivative processes for the mean function using the directional derivatives associated with the latent Gaussian process (Majumdar et al., 2006). Implementing such gradient analyses can be imagined for several spatial GLM frameworks, a few of which are outlined here. As before, we will assume that the covariate variable can additionally be assumed to follow a Gaussian process model.

The inference gained through post model fitting gradient analysis will have a slightly different nature for the spatial GLMs than for the standard spatial linear models. For a spatial linear model, say $Y(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s})$, we can investigate the local rate of change for the response $Y(\mathbf{s})$ at any location $\mathbf{s}$ and in any direction $\mathbf{u}$ through the directional derivative $D_{\mathbf{u}}Y(\mathbf{s})$. However, for spatial GLMs the response is typically non-continuous and a local rate of change is not meaningful. Taking binary data as an example, the responses will always be either 1 or 0 and there will never be a continuous change between responses. Instead we will consider the rates of change of a latent mean surface $\eta(\mathbf{s}) = g^{-1}(X(\mathbf{s})\beta)$, connected to the covariates through a link function. Unlike the non-continuous response surface, this latent mean surface will have a well defined directional derivative process which will inform on its rates of change.

Understanding the relationship between the latent mean surface and the covariate will be key to understanding the relationship with the response surface, though the precise interpretation will differ depending on the mean function η (e.g., probabilities vs. mean counts). The standard setup of the GLM allows for inference on this relationship at a global scale through the coefficient on the covariate in the linear form, β_1. We can further investigate variations in magnitude and strength of this relationship through the directional derivative processes. Examination of the local

directional sensitivity process can highlight local variation in the relationship both in terms of sign and magnitude along different directional perspectives by comparing the relative rates of change at each location in a given direction. In addition, the spatial angular discrepancy process will inform on the spatial variation in the strength of the relationship by comparing the directions of maximum gradient; a value of 0 suggests a strong positive relationship while a value of 2 suggests a strong negative relationship. This process is indicative of relationship strength since a pair of surfaces that are most rapidly increasing in the same or opposite direction throughout the region can be considered to have a stronger relationship than surfaces whose directions of maximum gradient are less consistent with one another.

There are two common frameworks for constructing a spatial GLM, and the choice of framework will be shown to impact the directional rates of change. The first framework incorporates a latent Gaussian process as a spatial random effect in the GLM's linear second stage. The second framework defines the response variable based on thresholding a latent Gaussian process. Both frameworks ultimately relate the mean surface to the covariates through a link function, but not surprisingly the different constructions lead to slightly different models and different gradient processes as a result.

For each of the spatial GLM examples we will assume the latent Gaussian process $Z(\mathbf{s})$ is centered at a linear function of a spatial covariate $X(\mathbf{s})$ which can be jointly modeled using a Gaussian process. I.e., in each model we assume the following relationship between $Z(\mathbf{s})$ and $X(\mathbf{s})$:

$$Z(\mathbf{s})|X(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_z(\mathbf{s})$$

$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$

$$w_z(\mathbf{s}) \sim GP(0, \rho_z)$$

$$w_x(\mathbf{s}) \sim GP(0, \rho_x)$$

where $w_z(\mathbf{s})$ and $w_x(\mathbf{s})$ are zero-mean spatial random effects surfaces.

To illustrate the first framework using spatial random effects consider binary and count data, two common non-Gaussian responses. For binary data either a spatial probit or logistic regression will typically be used where the observed presence or absence, $Y(\mathbf{s})$, is driven by the mean function $\eta(\mathbf{s}) = P(\mathbf{s}) = P(Y(\mathbf{s}) = 1)$. This latent probability surface relates to a Gaussian process $Z(\mathbf{s})$ through a link function, either $\Phi^{-1}(P(\mathbf{s})) = Z(\mathbf{s})$ for the probit regression or $logit(P(\mathbf{s})) = Z(\mathbf{s})$ for the logistic regression. Similarly, for count data one would use a spatial Poisson regression where the observed counts follow a Poisson distribution with mean $\eta(\mathbf{s}) = \lambda(\mathbf{s})$. A link function is again used to relate the mean surface to a latent Gaussian process, $\log(\lambda(\mathbf{s})) = Z(\mathbf{s})$.

The relevant local directional sensitivity process will relate the rate of change in the probability surface (for binary regression) or the mean surface (for Poisson regression) to the rate of change in the covariate surface. Through the spatial gradient chain rule these ratios can be expressed as simple functions of the independent Gaussian processes. For example, for the probit model the local directional sensitivity will be

$$\begin{aligned}\frac{D_{\mathbf{u}}P(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} &= \frac{D_{\mathbf{u}}\Phi(Z(\mathbf{s}))}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \frac{\phi(Z(\mathbf{s}))D_{\mathbf{u}}Z(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \phi(Z(\mathbf{s}))(\beta_1 + \frac{D_{\mathbf{u}}w_z(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})})\end{aligned}$$

at any location $\mathbf{s}$ and for any direction $\mathbf{u}$. I.e., the sensitivity for the latent probability surface can be written as the sensitivity for the latent Gaussian process, $D_{\mathbf{u}}Z(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, weighted by $\phi(Z(\mathbf{s}))$, the normal pdf for that surface at the loca-

tion $\mathbf{s}$. For Poisson regression the sensitivity will be

$$\begin{aligned}\frac{D_{\mathbf{u}}\lambda(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} &= \frac{D_{\mathbf{u}}\exp(Z(\mathbf{s}))}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \frac{\exp(Z(\mathbf{s}))D_{\mathbf{u}}Z(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \exp(Z(\mathbf{s}))(\beta_1 + \frac{D_{\mathbf{u}}w_z(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})})\end{aligned}$$

at any location $\mathbf{s}$ and for any direction $\mathbf{u}$. In this case, the sensitivity for the latent Gaussian process, $D_{\mathbf{u}}Z(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, is weighted by $\exp(Z(\mathbf{s}))$.

The second spatial GLM framework involves thresholding latent Gaussian processes. Bayesian modeling of non-Gaussian responses by thresholding latent normal random variables was originally proposed by Albert and Chib (1993), including guidelines for both binary and multi-category data. These ideas have been extended to accommodate spatially correlated data through the use of latent Gaussian processes in place of the independent normal random variables (Heagerty and Lele, 1998; Berrett and Calder, 2012). The mean function for the Gaussian process is assumed to be a regression on covariates, and as before the spatial dependence in the latent process encourages proximate locations to have similar presence probabilities.

For binary data the regression can be written

$$\begin{aligned}Y(\mathbf{s}) &= \begin{cases}1, & \text{if } Z(\mathbf{s}) > 0 \\ 0, & \text{if } Z(\mathbf{s}) < 0\end{cases} \\ Z(\mathbf{s}) &\sim GP(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}, \rho_z(1, \phi_z))\end{aligned} \tag{3.2}$$

where the probability of presence at any location $\mathbf{s}$ can be written

$$\begin{aligned}P(\mathbf{s}) &= P(Y(\mathbf{s}) = 1) \\ &= \Phi(\mathbf{X}(\mathbf{s})\boldsymbol{\beta})\end{aligned}$$

with $\Phi(\cdot)$ denoting the standard normal cdf and $\rho_z(\cdot)$ a spatial correlation function for the Gaussian process with the variance component set to 1 for identifiability.

To accommodate ordered multinomial data additional thresholds can be incorporated into the model in (3.2) such that each response category corresponds to a specific range of values for the latent variable (Albert and Chib, 1993). To induce spatial dependence, this model is again built using a Gaussian process specification. For a multinomial response $Y(\mathbf{s}) \in \{1, 2, 3\}$ we have

$$Y(\mathbf{s}) = \begin{cases} 2, & \text{if } a_1 < Z(\mathbf{s}) < a_2 \\ 1, & \text{if } a_0 < Z(\mathbf{s}) < a_1 \\ 0, & \text{if } a_{-1} < Z(\mathbf{s}) < a_0 \end{cases} \tag{3.3}$$

$$Z(\mathbf{s}) \sim GP(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}, \rho_z(1, \phi_z))$$

where $a_2 = \infty$ and $a_{-1} = -\infty$. The probability that category i will be observed at location $\mathbf{s}$ can be written

$$\begin{aligned} P_i(\mathbf{s}) &= P(Y(\mathbf{s}) = i) \\ &= \Phi(a_i - \mathbf{X}(\mathbf{s})\boldsymbol{\beta}) - \Phi(a_{i-1} - \mathbf{X}(\mathbf{s})\boldsymbol{\beta}) \end{aligned}$$

A similar model can be used to incorporate a point mass at zero, letting the response equal zero if the latent Gaussian process is negative and equal to the latent process otherwise:

$$Y(\mathbf{s}) = \begin{cases} Z(\mathbf{s}), & \text{if } Z(\mathbf{s}) > 0 \\ 0, & \text{if } Z(\mathbf{s}) < 0 \end{cases} \tag{3.4}$$

$$Z(\mathbf{s}) \sim GP(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}, \rho_z(1, \phi_z))$$

where the probability of a zero response is

$$\begin{aligned} P_0(\mathbf{s}) &= P(Y(\mathbf{s}) = 0) \\ &= 1 - \Phi(\mathbf{X}(\mathbf{s})\boldsymbol{\beta}) = \Phi(-\mathbf{X}(\mathbf{s})\boldsymbol{\beta}) \end{aligned}$$

The gradient analyses for these models consider the rate of change of the associated probability surface. Assuming a joint Gaussian process model for the latent

$Z(\mathbf{s})$ process and the covariate $X(\mathbf{s})$ process(es), the relationship between the probability surface and $X(\mathbf{s})$ involves a differentiable function and the spatial gradient chain rule can be applied in the usual way. For each of the above examples let the mean function be

$$\mathbf{X}(\mathbf{s})\boldsymbol{\beta} = \beta_0 + \beta_1 X(\mathbf{s})$$

$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$

$$w_x(\mathbf{s}) \sim GP(0, \rho_x(\sigma_x^2, \phi_x))$$

Then, applying the spatial gradient chain rule, we can compute the directional derivatives associated with each of the probability functions

$$D_{\mathbf{u}}P(\mathbf{s}) = \phi(\beta_0 + \beta_1 X(\mathbf{s}))\beta_1 D_{\mathbf{u}} w_x(\mathbf{s})$$

$$D_{\mathbf{u}}P_i(\mathbf{s}) = (\phi(a_{i-1} - \beta_0 - \beta_1 X(\mathbf{s})) - \phi(a_i - \beta_0 - \beta_1 X(\mathbf{s})))\beta_1 D_{\mathbf{u}} w_x(\mathbf{s})$$

$$D_{\mathbf{u}}P_0(\mathbf{s}) = -\phi(-\beta_0 - \beta_1 X(\mathbf{s}))\beta_1 D_{\mathbf{u}} w_x(\mathbf{s})$$

for the binary, multinomial and point mass at zero models respectively. Unlike the spatial random effects framework, the directional derivative processes for these probability surfaces do not involve the directional derivative for the latent Gaussian process $Z(\mathbf{s})$.

To compare the relative rates of change between the probability surface and the covariate surface the local directional sensitivity processes can be computed for each model

$$\frac{D_{\mathbf{u}}P(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \phi(\beta_0 + \beta_1 X(\mathbf{s}))\beta_1 \tag{3.5}$$

$$\frac{D_{\mathbf{u}}P_i(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = (\phi(a_{i-1} - \beta_0 - \beta_1 X(\mathbf{s})) - \phi(a_i - \beta_0 - \beta_1 X(\mathbf{s})))\beta_1$$

$$\frac{D_{\mathbf{u}}P_0(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = -\phi(-\beta_0 - \beta_1 X(\mathbf{s}))\beta_1$$

These local directional sensitivities do not involve the directional derivative processes for the $X(\mathbf{s})$ process nor the latent $Z(\mathbf{s})$ process, instead depending only on the

thresholds, $\boldsymbol{\beta}$ parameters, and the covariate process $X(\mathbf{s})$. Additionally, these forms reveal a lack of directional dependence. The directional derivative at location $\mathbf{s}$ will be the same for any choice of $\mathbf{u}$.

Alternatively, following the form for the spatial probit model described by Heagerty and Lele (1998), one can include a nugget effect so that the latent $Z(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + w_z(\mathbf{s}) + \epsilon(\mathbf{s})$ where $w_z(\mathbf{s}) \sim GP(0, \rho_z(\sigma_z^2, \phi_z))$ and $\epsilon(\mathbf{s}) \sim N(0, 1-\sigma_z^2)$. For the binary model, the resulting probability surface, associated directional derivative and local directional sensitivity will be

$$P(\mathbf{s}) = \Phi(\frac{\mathbf{X}(\mathbf{s})\boldsymbol{\beta} + w_z(\mathbf{s})}{\sqrt{1-\sigma_z^2}})$$

$$D_{\mathbf{u}}P(\mathbf{s}) = \phi(\frac{\mathbf{X}(\mathbf{s})\boldsymbol{\beta} + w_z(\mathbf{s})}{\sqrt{1-\sigma_z^2}})(\frac{\beta_1}{\sqrt{1-\sigma_z^2}}D_{\mathbf{u}}w_x(\mathbf{s}) + \frac{1}{\sqrt{1-\sigma_z^2}}D_{\mathbf{u}}w_z(\mathbf{s}))$$

$$\frac{D_{\mathbf{u}}P(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \phi(\frac{\mathbf{X}(\mathbf{s})\boldsymbol{\beta} + w_z(\mathbf{s})}{\sqrt{1-\sigma_z^2}})(\frac{\beta_1}{\sqrt{1-\sigma_z^2}} + \frac{1}{\sqrt{1-\sigma_z^2}}\frac{D_{\mathbf{u}}w_z(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})})$$

Unlike the local directional sensitivities in (3.5), when the nugget is included in the model the local directional sensitivity will depend on the relative rates of change of the latent Gaussian processes for $Z(\mathbf{s})$ and $X(\mathbf{s})$.

For each spatial GLM the spatial angular discrepancy process informs on the strength of the relationship between the mean surface and the covariate. It can be computed as $disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_\eta(\mathbf{s}))$, where $\theta_X(\mathbf{s})$ and $\theta_\eta(\mathbf{s})$ are the direction of maximum gradient for the covariate surface and mean surface respectively. These directions are easily computed as the angles associated with the unit vectors $\nabla_X(\mathbf{s})/||\nabla_X(\mathbf{s})||$ and $\nabla_\eta(\mathbf{s})/||\nabla_\eta(\mathbf{s})||$ shown to define the direction of maximum gradient by Banerjee et al. (2003).

3.4 South Africa Protea, Elevation Analysis

3.4.1 Cape Floristic Region Data

To demonstrate a binomial spatial GLM we consider a subset of the Protea Atlas data set (Rebelo, 2002) with elevation data from the South African Atlas of Hydrology and Climatology (Schultze, 1997) previously prepared for analysis by Gelfand et al. (2006) to examine species distributions and biodiversity. The data come from the Cape Floristic Region (CFR) in South Africa, the smallest of the world's six floral kingdoms (Takhtajan et al., 1986) and a well known biodiversity hotspot. The iconic flowering plant family in this region is the Protea family (Proteaceae), displaying high rates of endemism with 330 of the roughly 400 African species being 99% restricted to the CFR (Gelfand et al., 2006).

The subregion being considered is roughly rectangular with a total area of 4,456 km^2 and overlaps the Kogelberg Biosphere Reserve and surroundings. The data were originally collected at geo-referenced plots throughout the region, recording the presence or absence of protea species within each plot. Here, the data are provided in the form of 1554 minute by minute areal units (approximately 1.85 km $\times$ 1.55 km), aggregating information from all of the plots within each areal unit. For our purposes areal units that did not contain any plots provide no information and were removed from the data. This left a total of 962 areal units, with centroids plotted in Figure 3.1.

We focus on the protea species spear-leaved conebush (*Leucadendron spissifolium*), present in 33.7% of the areal units. For each areal unit the data provide a number of plots $N(\mathbf{s}) > 0$ visited within the cell, the number of plots $Y(\mathbf{s})$ where presence was observed, and an elevation $X(\mathbf{s})$. Interpolated surfaces for the observed presence rates $Y(\mathbf{s})/N(\mathbf{s})$ and elevation $X(\mathbf{s})$ are provided in Figure 3.1.

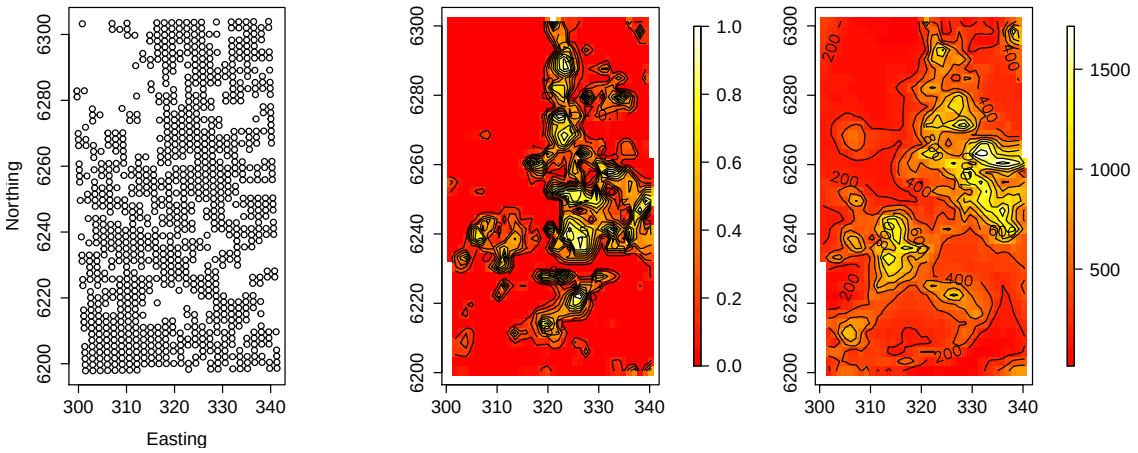

FIGURE 3.1: Centroids of observed areal units (left), interpolated $Y(\mathbf{s})/N(\mathbf{s})$ surface (center) and interpolated $X(\mathbf{s})$ surface (right).

3.4.2 Binomial Model and Spatial Gradients

These data are modeled using a spatial binomial regression with a logit link function, a natural extension of the binary models discussed earlier. For each areal unit centered at location $\mathbf{s}$ we observe a number of samples $N(\mathbf{s})$, with a total of $Y(\mathbf{s})$ presences observed for the given species, assumed to follow a binomial first stage model $[Y(\mathbf{s})|X(\mathbf{s})]$. To define our joint spatial model we implement a spatially varying intercept model for the elevation covariate $X(\mathbf{s})$.

For the areal unit centered at $\mathbf{s}$, let $P(\mathbf{s})$ be the presence probability and $Z(\mathbf{s})$ be a latent Gaussian process driving the presence probabilities. Then our model is

$$Y(\mathbf{s})|N(\mathbf{s}), P(\mathbf{s}) \sim Bin(N(\mathbf{s}), P(\mathbf{s}))$$

$$\log(\frac{P(\mathbf{s})}{1-P(\mathbf{s})}) = Z(\mathbf{s})$$

$$Z(\mathbf{s})|X(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_z(\mathbf{s})$$

$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$

$$w_z(\mathbf{s}) \sim GP(0, \rho_z(\phi_z, \sigma_z^2))$$

$$w_x(\mathbf{s}) \sim GP(0, \rho_x(\phi_x, \sigma_x^2))$$

where $w_z(\mathbf{s})$ and $w_x(\mathbf{s})$ represent independent spatial random effect processes associated with $Z(\mathbf{s})$ and $X(\mathbf{s})$ respectively. To satisfy required smoothness properties ensuring the existence of the associated directional derivative processes, the covariance functions $\rho_z(\cdot)$ and $\rho_x(\cdot)$ are set to be Matérn with smoothness parameter $\nu = 3/2$.

The above model is fit using the `'spBayes'` package in `R` (Finley et al., 2007) based on 5000 parameter samples. The $[Y(\mathbf{s})|X(\mathbf{s})]$ model has a burn-in of 1000 iterations and a thinning of every 15th iterate, while the $[X(\mathbf{s})]$ model has a burn-in of 1000 iterations and a thinning of every 5th iterate. Because the Gaussian first stage in the $X(\mathbf{s})$ model allows for marginalization over the $w_x(\mathbf{s})$ process during model fitting, there is reduced autocorrelation in the chains, suggesting the lower rate of thinning.

Note that these data are provided based on a grid laid over the region and are not available as truly point-referenced data. The binomial model will produce an estimated presence probability for each location $\mathbf{s}$, indexing an areal unit. The data format might suggest a preference for an areal model (e.g. a conditionally autoregressive (CAR) model), however, such an areal treatment of the data would preclude any analysis of rates of change for the presence probability surface. Instead, we opt for a Gaussian process model and assume the data observed at the grid cell level will inform about the presence probability at the central location $\mathbf{s}$, which can then be kriged in the usual way. These probabilities can be interpreted as describing the suitability of the location $\mathbf{s}$ to the species of interest.

The directional derivative processes associated with $Z(\mathbf{s})$ and $X(\mathbf{s})$ can be easily computed using samples of the directional derivatives, $D_{\mathbf{u}}w_x(\mathbf{s})$ and $D_{\mathbf{u}}w_z(\mathbf{s})$, associated with the independent Gaussian processes. Sampling these processes is straight forward, and simple functions of these will enable us to compute samples of $D_{\mathbf{u}}Z(\mathbf{s})$ and $D_{\mathbf{u}}X(\mathbf{s})$ for comparison using the local directional sensitivity process and the spatial angular discrepancy process.

The directional derivative processes associated with the probability surface $P(\mathbf{s}) = P(Y(\mathbf{s}) = 1)$ can then be computed by applying the spatial gradient chain rule to the inverse link function:

$$P(\mathbf{s}) = \frac{1}{1 + \exp(-Z(\mathbf{s}))}$$

$$D_{\mathbf{u}}P(\mathbf{s}) = \frac{\exp(-Z(\mathbf{s}))}{(1 + \exp(-Z(\mathbf{s})))^2} D_{\mathbf{u}}Z(\mathbf{s})$$

The resulting ratio of directional derivatives will then be

$$\begin{aligned} \frac{D_{\mathbf{u}}P(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} &= \frac{\exp(-Z(\mathbf{s}))}{(1 + \exp(-Z(\mathbf{s})))^2} \frac{D_{\mathbf{u}}Z(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \frac{\exp(-Z(\mathbf{s}))}{(1 + \exp(-Z(\mathbf{s})))^2} (\beta_1 + \frac{D_{\mathbf{u}}w_z(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})}) \end{aligned} \tag{3.6}$$

where the $D_{\mathbf{u}}Z(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ process is a spatial Cauchy process as described in Chapter 2, centered at the global relationship parameter β_1. This ratio corresponds to the local directional sensitivity process relating the latent probability surface to elevation. Again, the independence between the $w_z(\mathbf{s})$ and $w_x(\mathbf{s})$ processes enables convenient computation of the $D_{\mathbf{u}}P(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ process: one can first draw samples of the $D_{\mathbf{u}}w_z(\mathbf{s})$ and $D_{\mathbf{u}}w_x(\mathbf{s})$ processes independently, then utilize the expression given in (3.6) to compute the sensitivity.

Similarly, the relevant spatial angular discrepancy process can be defined as $disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_P(\mathbf{s}))$ where $\theta_X(\mathbf{s})$ and $\theta_P(\mathbf{s})$ are the directions of maximum gradient at location $\mathbf{s}$ for the covariate and presence probability surfaces respectively. This process informs on the strength of the relationship between presence probability and elevation across the region, with 0 or 2 indicating a strong positive or negative relationship and values in between suggesting a weaker relationship.

3.4.3 Results

Table 3.1 provides fitted parameter values and 95% credible intervals for the $[Y(\mathbf{s})|X(\mathbf{s})][X(\mathbf{s})]$ model. The coefficient on altitude, β_1, is significantly positive, suggesting that higher altitudes are associated with higher odds of presence. However, since the models are generalizing over the entire region, averaging over local variation, this parameter is a fairly rough estimate of the relationship between the variables.

Under this model there is a latent presence probability surface $P(\mathbf{s})$ being driven by the $Z(\mathbf{s})$ Gaussian process. The posterior mean $P(\mathbf{s})$ surface is provided in Figure 3.2 which generally shows higher presence probabilities with higher altitudes, in alignment with the interpretation of β_1.

Table 3.1: Parameter estimates for the $[Y(\mathbf{s})|X(\mathbf{s})][X(\mathbf{s})]$ model.

Parameter	0.025	Mean	0.975
β_0	-4.9335	-4.2571	-3.5649
β_1	0.0011	0.0019	0.0027
α_0	347.8432	411.3375	472.5418
ϕ_z	0.5590	0.6763	0.8000
σ_z^2	4.3078	5.7106	7.5332
ϕ_x	0.5994	0.6501	0.7003
σ_x^2	91378.8	106727.2	126015.6

To learn about the variables' relationship at a more local scale we turn to the gradient processes developed in Chapter 2. Figure 3.3 provides the posterior median local directional sensitivity process $D_{\mathbf{u}}P(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ looking north/south and east/west, and Figure 3.4 provides the posterior median spatial angular discrepancy process, $disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_P(\mathbf{s}))$. Recalling that a discrepancy of 2 indicates a strong negative relationship and a discrepancy of 0 indicates a strong positive relationship, it appears the three surfaces are providing similar insight into the relationship between the variables.

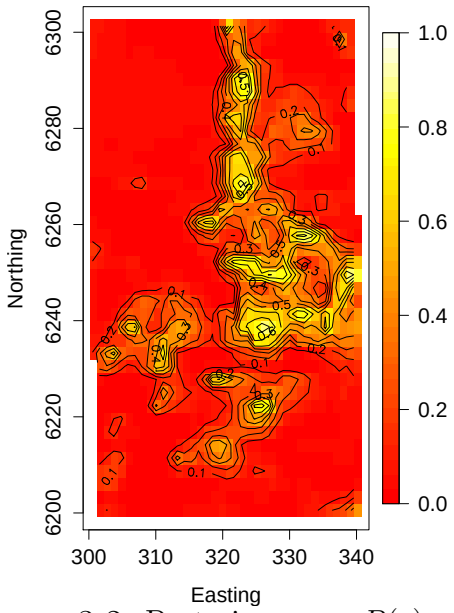

FIGURE 3.2: Posterior mean $P(\mathbf{s})$ surface.

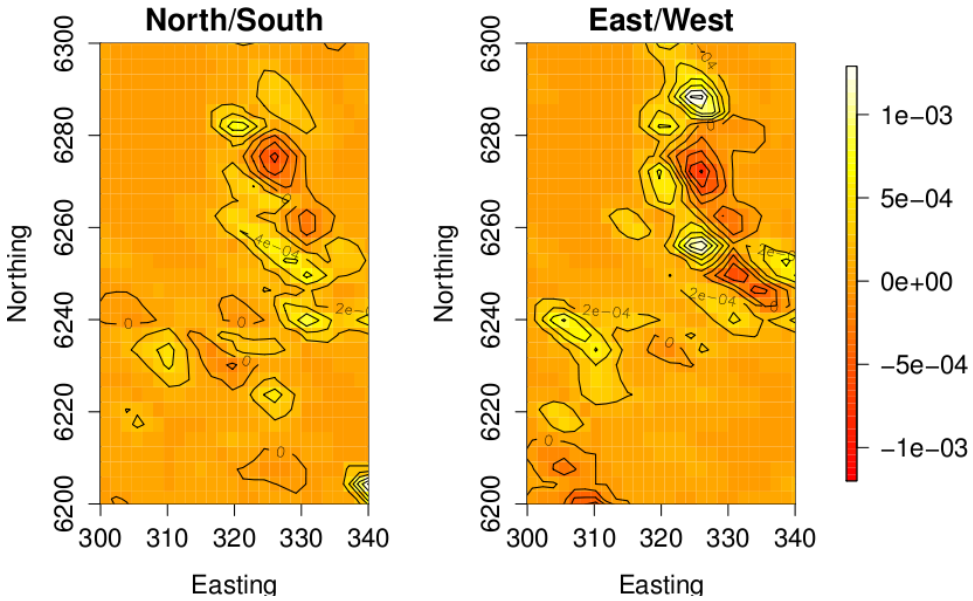

FIGURE 3.3: Posterior median $D_{\mathbf{u}}P(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ for $\mathbf{u}$ North/South (left) and East/West (right).

When compared to the altitude surface provided in Figure 3.1, the sensitivity and discrepancy processes suggest that the relationship between presence probability and altitude varies based on altitude. The areas showing a negative relationship generally

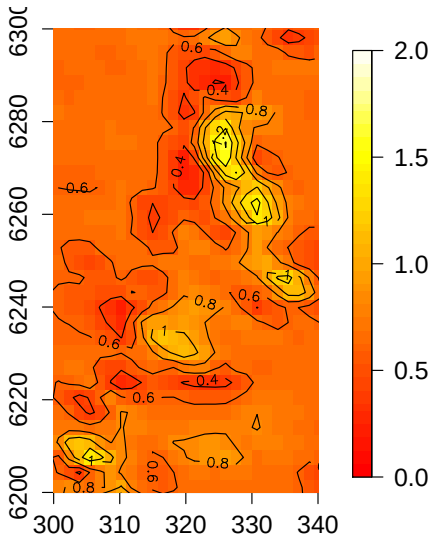

FIGURE 3.4: Posterior median $disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_P(\mathbf{s}))$.

correspond to the areas with the highest elevation ($>\sim 900$); the areas with a positive relationship correspond to areas with middling elevation ($\sim 200 - 900$); and the areas with a neutral relationship correspond to areas with low elevation ($<\sim 200$). Biologically, this would suggest that the species cannot survive at low elevations so no relationship is observed. Once the altitude reaches a certain level (~ 200) the species starts to occur, and its presence increases as the altitude increases. Finally, if the altitude exceeds some tolerance level (~ 900) the species will start struggling to survive and species presence will decrease as the altitude continues to increase. Such a pattern can be observed through the sensitivity and discrepancy processes, but would otherwise be difficult to detect using this model.

3.5 Duke Forest Point Pattern, Elevation Analysis

3.5.1 Duke Forest Data

To illustrate a more novel non-Gaussian dataset we turn to a collection of point patterns of tree species present at the Blackwood site in the Duke Forest in Durham, NC. The site is 5 hectares in area and exhibits a range of elevation. A road and

powerline separate the site into three subregions, and we focus on the northwestern of these subregions. We consider two tree species, Flowering Dogwood (*Cornus florida*) and Sweetgum (*Liquidambar styraciflua*), with regard to their respective point patterns of locations within the site in the year 2000. Some trees have multiple stems observed at a single location; however we treat these as a single observed tree at the given location. The point pattern for Sweetgum consists of 531 trees, and the point pattern for Flowering Dogwood consists of 570 trees. Elevation is recorded at each location where a tree of any species was observed, resulting in 5654 elevation observations.

Figure 3.5 provides an interpolation of the elevation data. There is a clear increase in elevation across the region in a roughly southeastern direction. Figure 3.5 also provides the observed point patterns for each of the species. Flowering Dogwood is well dispersed across the entire region while Sweetgum is more abundant in the northwestern half of the region.

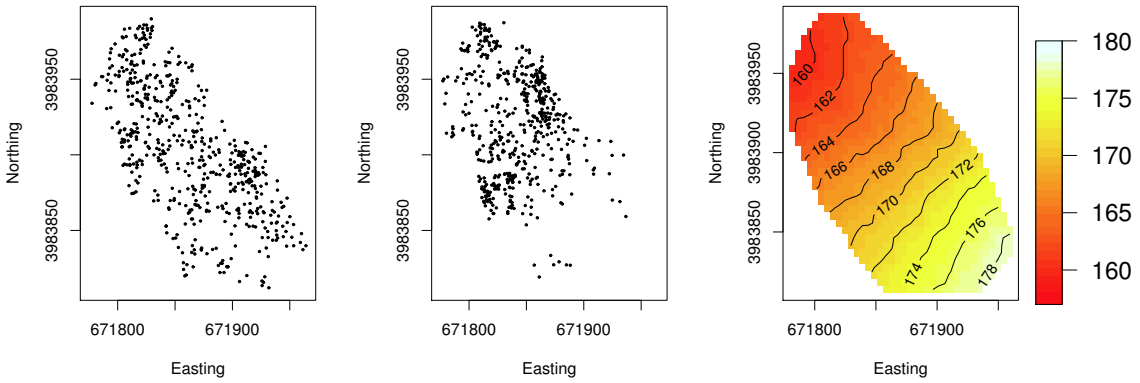

FIGURE 3.5: From left to right, observed point patterns for Flowering Dogwood and Sweetgum and observed elevation.

3.5.2 Point Pattern Model and Spatial Gradients

Given the observed elevation, $X(\mathbf{s})$, we model the intensity for each species as a log-Gaussian Cox process: $\lambda(\mathbf{s}) = \exp(\beta_0 + \beta_1 X(\mathbf{s}))\lambda_0(\mathbf{s})$, $\lambda_0(\mathbf{s}) = \exp(w_z(\mathbf{s}))$, and

$w_z(\mathbf{s}) \sim GP(0, \rho(\cdot|\phi_z, \sigma_z^2))$. Placing priors on the parameters, we have the following model:

$$[\tilde{Y}|\beta_0, \beta_1, \{\lambda_0(\mathbf{s}), \mathbf{s} \in D\}][\lambda_0(\mathbf{s})|\sigma^2, \phi][\beta_0, \beta_1][\phi][\sigma^2]$$
$$= [\tilde{Y}|\beta_0, \beta_1, \{w_z(\mathbf{s}), \mathbf{s} \in D\}][w_z(\mathbf{s})|\sigma^2, \phi][\beta_0, \beta_1][\phi][\sigma^2]$$

where $\tilde{Y}$ is the observed point pattern and $\lambda_0(\mathbf{s})$ is equivalently considered using $w_z(\mathbf{s})$.

We approximate the likelihood by dividing the region into a fine grid with cells $\{A_l; l = 1, \ldots, L\}$. This gives us the likelihood

$$L(\lambda(\mathbf{s}), \mathbf{s} \in D; \mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n) \approx \Pi_i \lambda(\mathbf{s}_i) \exp(-\lambda(D))$$
$$\lambda(D) \approx \Sigma_l \exp(X'(A_l)\beta + w_z(A_l))$$

where $w_z(A_l)$ corresponds to a realization from a Gaussian process evaluated at a representative point in each grid cell A_l. This likelihood can be sampled using elliptical slice sampling, as described in Murray et al. (2009) and Murray and Adams (2010).

In terms of $Z(\mathbf{s}) = \log(\lambda(\mathbf{s}))$, we immediately have a conditional bivariate GP model with elevation: $Z(\mathbf{s})|X(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_z(\mathbf{s})$ and $X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$ with $w_z(\mathbf{s}) \sim GP(0, \rho(\cdot|\sigma_z^2, \phi_z))$ and $w_x(\mathbf{s}) \sim GP(0, \rho(\cdot|\sigma_x^2, \phi_x))$. Hence, we are in the framework developed in Chapter 2 and can apply the proposed gradient analyses. The difference in this case will be that the Gaussian response $Z(\mathbf{s})$ is latent and thus unobserved. The uncertainty about $Z(\mathbf{s})$ is propagated through the model by drawing a posterior sample of the $Z(\mathbf{s})$ surface for each posterior sample of the parameters.

The methods developed in Chapter 2 allow for straight forward examination of $D_{\mathbf{u}}\log\lambda(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s}) = D_{\mathbf{u}}Z(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, although interest is more likely in the behavior of the intensity surface itself. Applying the spatial gradient chain rule

discussed in Section 3.2 gives us $D_{\mathbf{u}}\lambda(\mathbf{s}) = \exp(Z(\mathbf{s}))D_{\mathbf{u}}Z(\mathbf{s})$, with the directional derivative ratio $D_{\mathbf{u}}\lambda(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s}) = \exp(Z(\mathbf{s}))D_{\mathbf{u}}Z(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$. As before, we can simplify this in terms of the independent Gaussian processes:

$$\begin{aligned}\frac{D_{\mathbf{u}}\lambda(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} &= \exp(Z(\mathbf{s}))\frac{D_{\mathbf{u}}Z(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} \\ &= \frac{\exp(Z(\mathbf{s}))[\beta_1 D_{\mathbf{u}}w_x(\mathbf{s}) + D_{\mathbf{u}}w_z(\mathbf{s})]}{D_{\mathbf{u}}w_x(\mathbf{s})} \\ &= \exp(Z(\mathbf{s}))[\beta_1 + \frac{D_{\mathbf{u}}w_z(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})}]\end{aligned}$$

The Cauchy random variable will again be centered at β_1, but now there will also be scaling according to the value of $\exp(Z(\mathbf{s}))$.

3.5.3 Results

Table 3.5.3 provides the fitted parameter values for elevation model. Tables 3.5.3 and 3.5.3 provide the fitted parameter values for each of the species models. Note that the ϕ_z parameter is fixed at the minimum contrast estimate, as suggested in Møller et al. (1998), to facilitate identifiability in the fitting of the log-Gaussian Cox process. The fixed values are $\phi_z = 0.1063$ and 0.0434 for Flowering Dogwood and Sweetgum respectively. The 95% credible interval for β_1 contains zero for Flowering Dogwood, but not for Sweetgum. This negative coefficient suggests that the intensity of Sweetgum decreases as elevation increases, while the intensity of Flowering Dogwood is not responsive to elevation changes at this scale. Figure 3.6 provides the posterior median intensity for each of the species. Both intensities have fairly low values across most of the domain, with a few regions of higher intensity.

In Figure 3.5 we saw a clear increase in elevation in a roughly southeastern direction. We approximate this direction by the unit vector $\mathbf{u} = (0.8508, -0.5255)$ and consider the behavior of the directional derivative ratios in this direction for

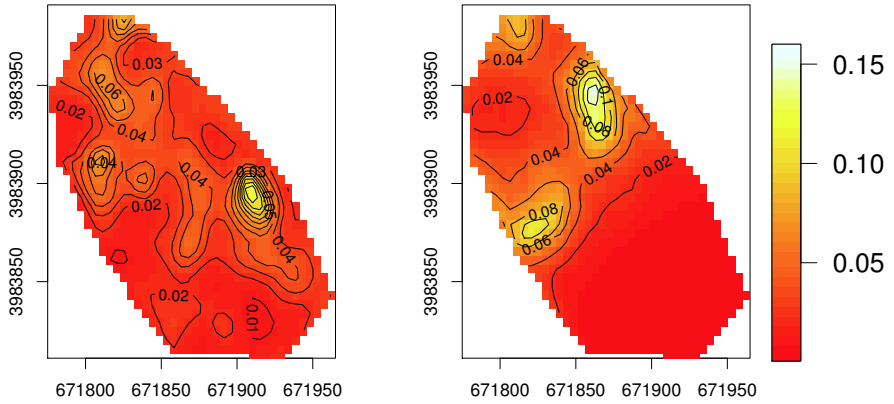

FIGURE 3.6: Posterior median of the intensity surface for Flowering Dogwood (left) and Sweetgum (right).

Table 3.2: Parameter estimates for $[X(\mathbf{s})]$ model for elevation.

Parameter	0.025	Mean	0.975
α_0	166.1043	167.8577	169.4328
σ_x^2	7.1681	9.4117	12.9494
ϕ_x	0.0791	0.0892	0.0979

each of the species. Figure 3.7 plots the resulting posterior median $D_{\mathbf{u}}\lambda(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ surfaces.

For Flowering Dogwood the majority of the domain has a ratio close to zero. This suggests that the changes in the intensity are negligible compared to the changes in elevation. Recalling the fairly even spread of the trees in the region, as well as the non-significant β_1, this pattern makes sense.

For Sweetgum, virtually the entire region has a negative directional derivative ratio. This aligns with our interpretation of the significantly negative β_1 coefficient, namely that as elevation increases the intensity decreases. There are a few subregions where the change in intensity occurs more rapidly than elsewhere, and there is a larger subregion where the change in intensity is zero due to an absence of trees. The cause for subregions of rapid change could be further illuminated through examination

Table 3.3: Parameter estimates for $[Z(\mathbf{s})|X(\mathbf{s})]$ model for Flowering Dogwood.

Parameter	0.025	Mean	0.975
β_0	-3.6416	-3.5478	-3.4590
β_1	-0.0866	-0.0284	0.0348
σ_z^2	0.3367	0.5237	0.7648

Table 3.4: Parameter estimates for $[Z(\mathbf{s})|X(\mathbf{s})]$ model for Sweetgum.

Parameter	0.025	Mean	0.975
β_0	-4.2161	-4.0297	-3.8728
β_1	-0.4094	-0.2619	-0.1024
σ_z^2	0.7942	1.2652	1.9160

of other factors in those regions. Similarly, the region of zero change in intensity could be roughly interpreted as Sweetgum having an aversion to elevations beyond a certain value.

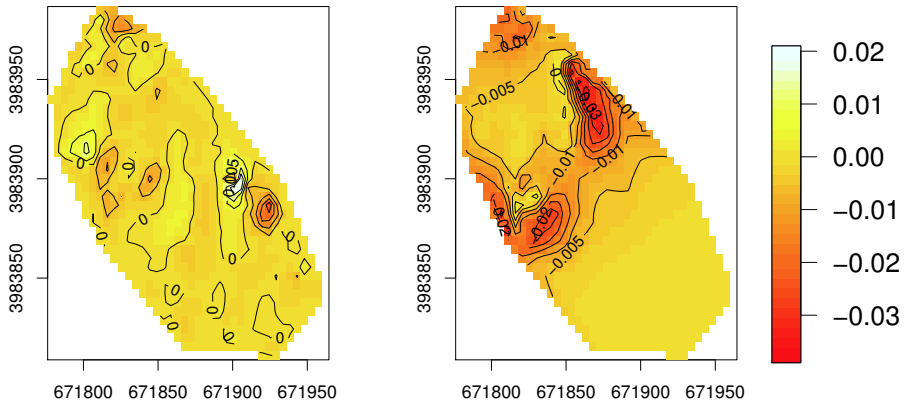

FIGURE 3.7: Posterior median of $D_{\mathbf{u}}\lambda(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ for Flowering Dogwood (left) and Sweetgum (right); $\mathbf{u} = (0.8508, -0.5255)$.

Finally, we can compare the intensity and elevation surfaces by computing the discrepancy between their directions of maximum gradient at each location, i.e., the posterior median of $disc(\mathbf{s}) = 1 - \cos(\theta_X(\mathbf{s}) - \theta_Y(\mathbf{s}))$ across the region. Values close to 2 suggest the surfaces are most rapidly increasing in opposite directions; values

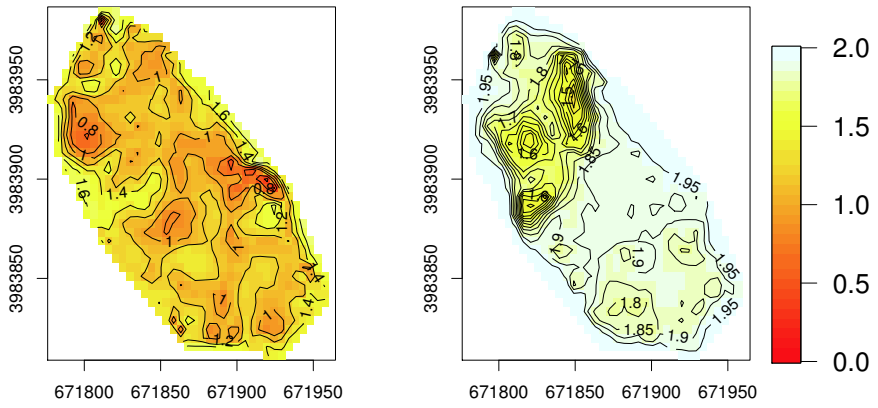

FIGURE 3.8: Posterior median $disc(\mathbf{s})$ for Flowering Dogwood (left) and Sweetgum (right).

close to 0 suggest the surfaces are most rapidly increasing in the same direction. The posterior median discrepancy surfaces are provided in Figure 3.8 for Flowering Dogwood and Sweetgum.

The discrepancies for Flowering Dogwood roughly range between 0.8 and 1.5. There is no clear pattern, which supports there being no strong relationship between Flowering Dogwood intensity and elevation.

The pattern for Sweetgum is quite different. All of the discrepancies appear to be between 1.5 and 2, with most around 1.9. This suggests that the Sweetgum intensity and elevation are increasing in nearly opposite directions virtually everywhere in the domain. This again confirms the negative relationship, and additionally highlights this pattern as being slightly weaker in the northern part of the region.

3.6 Summary

Accounting for spatial dependence in observations can benefit analyses for a wide array of real world data which, not surprisingly, will often require a non-Gaussian first stage model. Many of these models will link a latent Gaussian process to the

response surface of interest through a differentiable function. As shown here, it is straight forward to apply post model fitting gradient analysis in these cases through the use of the spatial gradient chain rule. The resulting gradient analysis procedure provides more localised inference regarding the relationship between the variables without additional model fitting.

Several model structures exist for non-Gaussian data, several of which are discussed for binary data by Paciorek (2007). Here we show that the model choice will have implications on the rates of change for the corresponding probability surfaces. As demonstrated for a spatial probit regression, when using a thresholding approach without a nugget term the rates of change will depend only on the rates of change of the covariate surface. If a nugget term is incorporated into the model then the rate of change will depend on both the rate of change for the latent $Z(\mathbf{s})$ and the covariate surface $X(\mathbf{s})$. Finally, under a link function specification, the rate of change will be similar to the thresholding model with a nugget, without the weighting of $1/\sqrt{1-\sigma_z^2}$. While similar, each model formulation suggests a distinct behavior for the rates of change that may or may not be desirable. Similar differences can be observed for other spatial GLMs.

To illustrate the gradient analysis in a spatial GLM framework we analyzed presence-absence data coming from the CFR in South Africa requiring a binomial first stage model. Standard inference procedures would involve examination of the coefficient β_1 to learn about the relationship with altitude as well as the spatial $P(\mathbf{s})$ surface to visualize the spatial distribution of the species. For these data, the researcher would conclude that the species is positively associated with higher altitudes. However, the local directional sensitivity process and spatial angular discrepancy process tell a slightly different story. By highlighting more localized relationship patterns these processes reveal that the positive relationship with altitude only occurs for certain levels of altitude, with the highest altitudes exhibiting a neg-

ative effect on species presence. Such conclusions regarding the range limits for the species could have impacts on management decisions in the CFR.

Next, we applied gradient analyses to data where the response of interest was a point pattern of trees from the Duke Forest modeled as a function of elevation using a log-Gaussian Cox process. Using the local directional sensitivity process and the spatial angular discrepancy process we studied the sensitivity of the intensity to changes in elevation for two tree species. Only one of the two species is responsive to changes in elevation, allowing for comparison of these gradient processes across significant and non-significant regressions.

4

Gradient Analysis for Multivariate Spatial Processes

4.1 Introduction

A common inferential aspect of spatial modeling is the resulting description of variable relationships and how those relationships vary across a region. Post model fitting gradient analyses can further illuminate these relationships, as described for bivariate data in Chapter 2 and non-Gaussian data in Chapter 3. However, it is well understood that no two variables will interact with one another in a vacuum, completely independent of their surroundings. Instead, there will typically be a variety of factors impacting the observed response at any location, and incorporating these additional factors into a given model will have the potential to enhance our understanding of these complex relationships. When these additional factors can reasonably be modeled as a spatial surface they should be incorporated into the joint spatial modeling, thus necessitating an extension of the gradient analysis procedures to accommodate multivariate spatial processes.

The methodology proposed here will enable gradient analyses to be applied to

data where there may be several spatial response variables and/or several spatial covariates of interest. In a bivariate example the only relationship that needs to be explored is the dependence between response and covariate. In contrast, a multivariate example requires consideration of all response-covariate pairs, dependencies which may all vary spatially in different ways. In addition, these pairwise dependencies can be considered either marginally or conditional on the other model variables. The gradient analyses provide further insight into these relationships as defined by the joint model.

To illustrate this methodology we explore a dataset relating plant leaf traits and climate in the Cape Floristic Region of South Africa. As described in Lamont et al. (2002), within a plant family variation in leaf morphology will reflect the variable environmental constraints on plant growth. Understanding these constraints and how they vary across the region can provide insight into how plants may respond to changes in climate. In our setup leaf traits are modeled jointly without assuming any explicit dependence, as are temperature and precipitation, and all variables are assumed to follow a spatial process model. This framework captures the correlation between leaf traits which are not inherently independent since different trait measurements are taken from a single individual. The dependence between leaf traits and climate is explored first through the model parameters and later through the gradient analyses.

The joint models and relevant gradient theory are detailed in Section 4.2; subsection 4.2.1 describes a single spatial response with multiple spatial covariates, subsection 4.2.2 describes multiple spatial responses with a single covariate, and subsection 4.2.3 describes multiple responses and multiple covariates. The data for the analyses are described in Section 4.3. Results for the former two scenarios are provided in Section 4.4. An overall summary is provided in Section 4.5.

4.2 Modeling Development

We consider the class of multivariate spatial coregionalization models and discuss methodology for conducting gradient analyses within each setup. In particular, we discuss three general frameworks for multivariate modeling: (1) one response with multiple covariates, (2) multiple response variables with one covariate, and (3) multiple responses and multiple covariates. Each model implies a distribution for the associated directional derivative processes which can be used to gain deeper insight into the spatial relationships between the variables.

The processes are assumed, for convenience, to be weakly stationary such that the covariance function $C(\mathbf{s}, \mathbf{s}')$ depends only on the separation vector $\boldsymbol{\delta} = \mathbf{s} - \mathbf{s}'$. Specifically, we focus on isotropic covariance functions that depend only on the length of the separation vector, $||\boldsymbol{\delta}||$, and in our data examples the covariance function will be assumed to be Matérn with smoothness parameter $\nu = 3/2$.

4.2.1 Multiple Predictors

Consider a response surface $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ and two correlated covariate surfaces $\{X_1(\mathbf{s}), X_2(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$. Let $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), X_2(\mathbf{s}))'$ and $\mathbf{X}(\mathbf{s}) = \boldsymbol{\gamma} + \mathbf{A}\mathbf{W}(\mathbf{s})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$, $\mathbf{A}$ is a symmetric 2×2 matrix and $\mathbf{W}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}))'$ is a pair of independent zero-mean Gaussian processes with spatial covariance functions $G_1(\boldsymbol{\delta}; 1, \phi_1)$ and $G_2(\boldsymbol{\delta}; 1, \phi_2)$ respectively. Note that the variance parameters in the covariance functions are set to 1 for identifiability. Model the response surface conditional on the covariates as a realization from a Gaussian process with covariance function $K(\boldsymbol{\delta}; \sigma_y^2, \phi_y)$ such that $E(Y(\mathbf{s})|\mathbf{X}(\mathbf{s})) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} = \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s})$. Observed at a set of locations $\mathbf{Y} = (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))$, we write $\mathbf{Y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{K})$ and $\mathbf{X} \sim N(\boldsymbol{\gamma}, \mathbf{G_1} \otimes \mathbf{A_1} + \mathbf{G_2} \otimes \mathbf{A_2})$, where $\mathbf{G}_i$ is the covariance matrix associated with w_i and $\mathbf{A}_i = \mathbf{a}_i\mathbf{a}_i'$ with $\mathbf{a}_i$ the ith column of $\mathbf{A}$.

Extending this model to accommodate additional predictor variables would be straight forward, requiring an additional zero-mean Gaussian process $w_i(\mathbf{s})$ for each additional variable. The general model for $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), X_2(\mathbf{s}), \ldots, X_m(\mathbf{s}))'$ would then be $\mathbf{X} \sim N(\boldsymbol{\gamma}, \Sigma_{i=1}^{m} \mathbf{G_i} \otimes \mathbf{A_i})$. Similarly, when appropriate, a differentiable function of one or more spatial covariates can be incorporated as an additional predictor. It is straightforward to extend the chain rule to accommodate a multivariate function of spatial surfaces, such as an interaction term. For some differentiable function $g(X_1(\mathbf{s}), X_2(\mathbf{s}))$ the associated directional derivative will be $D_{\mathbf{u}} g(X_1(\mathbf{s}), X_2(\mathbf{s})) = dg(X_1(\mathbf{s}), X_2(\mathbf{s}))/dX_1(\mathbf{s}) \times D_{\mathbf{u}} X_1(\mathbf{s}) + dg(X_1(\mathbf{s}), X_2(\mathbf{s}))/dX_2(\mathbf{s}) \times D_{\mathbf{u}} X_2(\mathbf{s})$. For example, an interaction between the covariates, $g(X_1(\mathbf{s}), X_2(\mathbf{s})) = X_1(\mathbf{s})X_2(\mathbf{s})$, would have a directional derivative $D_{\mathbf{u}}(X_1(\mathbf{s})X_2(\mathbf{s})) = X_1(\mathbf{s})D_{\mathbf{u}}X_2(\mathbf{s}) + X_2(\mathbf{s})D_{\mathbf{u}}X_1(\mathbf{s})$.

Prior distributions for the parameters completes the model specification. For the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters independent broad Normal priors are assumed. Inverse Gamma distributions are used for the range parameters in the covariance functions ϕ_1, ϕ_2 and ϕ_y, and similarly for the variance parameter σ_y^2 in the covariance function for $Y(\mathbf{s})$.

To understand the relationship between response and covariates there is interest in sampling the associated directional derivatives $D_{\mathbf{u}}Y(\mathbf{s})$, $D_{\mathbf{u}}X_1(\mathbf{s})$, and $D_{\mathbf{u}}X_2(\mathbf{s})$. In lieu of working with large joint multivariate Normal distributions, sampling these derivatives is made more computationally convenient by captializing on independent zero-mean spatial processes. With this goal, the above model can be written

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + w_y(\mathbf{s}) \\ X_1(\mathbf{s}) &= \gamma_1 + a_{11} w_1(\mathbf{s}) + a_{12} w_2(\mathbf{s}) \\ X_2(\mathbf{s}) &= \gamma_2 + a_{12} w_1(\mathbf{s}) + a_{22} w_2(\mathbf{s}) \end{aligned} \tag{4.1}$$

where $w_y(\mathbf{s})$, $w_1(\mathbf{s})$, and $w_2(\mathbf{s})$ are independent Gaussian processes. Given posterior samples of the model parameters it is straight forward to sample the directional derivative processes $D_{\mathbf{u}} w_i(\mathbf{s})$ marginally, thus avoiding any explosion in dimension.

Based on these samples, we can then compute the directional derivatives for the response and covariates of interest.

The forms for the local directional sensitivity processes $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$ and $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X_2(\mathbf{s})$ as functions of the independent Gaussian processes are provided in (4.2) and (4.3).

$$\begin{aligned}\frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} &= \beta_1 + \frac{\beta_2 D_{\mathbf{u}}X_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} \\ &= \beta_1 + \frac{\beta_2 a_{12} D_{\mathbf{u}}w_1(\mathbf{s}) + \beta_2 a_{22} D_{\mathbf{u}}w_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{a_{11} D_{\mathbf{u}}w_1(\mathbf{s}) + a_{12} D_{\mathbf{u}}w_2(\mathbf{s})} \qquad (4.2)\end{aligned}$$

$$\begin{aligned}\frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} &= \beta_2 + \frac{\beta_1 D_{\mathbf{u}}X_1(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} \\ &= \beta_2 + \frac{\beta_1 a_{11} D_{\mathbf{u}}w_1(\mathbf{s}) + \beta_1 a_{12} D_{\mathbf{u}}w_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{a_{12} D_{\mathbf{u}}w_1(\mathbf{s}) + a_{22} D_{\mathbf{u}}w_2(\mathbf{s})} \qquad (4.3)\end{aligned}$$

In contrast to the single-covariate model, the sensitivities here exhibit dependence in the ratio numerator and denominator. Understanding the distributional effects of this dependence proves illuminating regarding the interpretation of the processes.

Recall that the sensitivities associated with the single-covariate model were centered at the parameter controlling the global relationship between the two variables. A similar result can be obtained for the multi-covariate framework: If $Var(D_{\mathbf{u}}w_1(\mathbf{s})) = Var(D_{\mathbf{u}}w_2(\mathbf{s}))$, then the directional derivative ratios will be centered at the global coefficients for $X_2(\mathbf{s})$ and $X_1(\mathbf{s})$. If the variances are *not* equal, then the centering for the ratios will be 'close' to the global coefficient, as a function of the variance parameters. We define the global coefficient for $X_i(\mathbf{s})$ as $\beta_{i|x_i}$ such that $E(Y(\mathbf{s})|X_i(\mathbf{s})) = \beta_{0|x_i} + \beta_{i|x_i}X_i(\mathbf{s})$; that is, it is the effect of the covariate $X_i(\mathbf{s})$ after marginalizing over the other covariate(s).

To illustrate this centering for $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$, consider the conditional expec-

tation of $Y(\mathbf{s})|X_1(\mathbf{s})$ using standard multivariate normal theory:

$$E(Y(\mathbf{s})|X_1(\mathbf{s})) = \beta_0 + \beta_2\gamma_2 - \frac{\beta_2(a_{11}a_{21} + a_{22}a_{21})}{a_{11}^2 + a_{21}^2}\gamma_1 + (\beta_1 + \frac{\beta_2(a_{11}a_{21} + a_{22}a_{21})}{a_{11}^2 + a_{21}^2})X_1$$

$$= \beta_{0|x_1} + \beta_{1|x_1}X_1(\mathbf{s}) \tag{4.4}$$

To see that this is the centering for the sensitivity, rewrite the ratio as follows:

$$\begin{aligned}\frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} =& \frac{(\beta_1 a_{11} + \beta_2 a_{21})D_{\mathbf{u}}w_1(\mathbf{s}) + (\beta_1 a_{21} + \beta_2 a_{22})D_{\mathbf{u}}w_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{21}D_{\mathbf{u}}w_2(\mathbf{s})} \\ =& \frac{(\beta_1 a_{11} + \beta_2 a_{21})D_{\mathbf{u}}w_1(\mathbf{s})}{a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{21}D_{\mathbf{u}}w_2(\mathbf{s})} + \frac{(\beta_1 a_{21} + \beta_2 a_{22})D_{\mathbf{u}}w_2(\mathbf{s})}{a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{21}D_{\mathbf{u}}w_2(\mathbf{s})} \\ &+ \frac{D_{\mathbf{u}}w_y(\mathbf{s})}{a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{21}D_{\mathbf{u}}w_2(\mathbf{s})}\end{aligned} \tag{4.5}$$

The three ratios in (4.5) will be dependent Cauchy random variables. The third ratio is centered at 0, and it can be shown that the first two ratios will be centered at $(\beta_1 a_{11}/a_{21} + \beta_2)\frac{a_{11}/a_{21}}{a_{11}^2/a_{21}^2 + v_2^2/v_1^2}$ and $(\beta_1 a_{21}/a_{11} + \beta_2 a_{22}/a_{11})\frac{a_{21}/a_{11}}{a_{21}^2/a_{11}^2 + v_1^2/v_2^2}$ respectively, where $Var(D_{\mathbf{u}}w_i(\mathbf{s})) = v_i^2$. Combining these terms, the overall centering for (4.5) will be $\beta_1 + \beta_2\frac{v_1^2 a_{11}a_{21} + v_2^2 a_{22}a_{21}}{v_1^2 a_{11}^2 + v_2^2 a_{21}^2}$. If $v_1 = v_2$, this centering will be equal to $\beta_{1|x_1}$ defined in (4.4). An analogous result exists for X_2.

In the case of the Matérn covariance $Var(D_{\mathbf{u}}w_i(\mathbf{s})) = \phi_i^2$, so centering at the global coefficient will occur when the covariates share a common range parameter $\phi_1 = \phi_2$.

This result provides an intuitive interpretation for the local directional sensitivity processes. Parameter estimates from the model fitting will inform on the *conditional* effects of each covariate, after accounting for the additional covariate(s). In a complementary fashion, examination of the local directional sensitivities provides insight into the *marginal* effect of the covariate across spatial and directional perspectives.

A hierarchical alternative to the model in (4.1) was elaborated in Royle and Berliner (1999) and can be useful in scenarios where there is a natural ordering to

the variables. As described in Banerjee et al. (2004), the conditional multivariate model can be viewed as a coregionalization model, $\mathbf{X}(\mathbf{s}) = \boldsymbol{\gamma} + \mathbf{A}\mathbf{W}(\mathbf{s})$, but arises through specificication of $\mathbf{A}$ as an upper triangular matrix as opposed to the symmetric matrix described earlier. The general equivalence between the conditional and unconditional specifications was shown in Banerjee et al. (2004), however the conditional specification induces asymmetry in the local directional sensitivity processes that may be undesirable when the variables have no inherent ordering.

Analogous computations for the conditional coregionalization model produce similar inference and interpretations for the sensitivities. The model can be written,

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + w_y(\mathbf{s}) \\ X_2(\mathbf{s}) &= \alpha_0 + \alpha_1 X_1(\mathbf{s}) + \tilde{w}_2(\mathbf{s}) \\ X_1(\mathbf{s}) &= \eta_0 + \tilde{w}_1(\mathbf{s}) \end{aligned} \tag{4.6}$$

where $w_y(\mathbf{s})$, $\tilde{w}_1(\mathbf{s})$, and $\tilde{w}_2(\mathbf{s})$ are again independent Gaussian processes. Note, we write $\tilde{w}$ for notational purposes to differentiate the $\tilde{w}_1(\mathbf{s})$ and $\tilde{w}_2(\mathbf{s})$ processes from the $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$ processes.

Not surprisingly, the difference in model specification produces a difference in the forms for the local directional sensitivity processes $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$ and $D_{\mathbf{u}}Y(\mathbf{s})/D_{\mathbf{u}}X_2(\mathbf{s})$, provided in (4.7) and (4.8).

$$\begin{aligned} \frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} &= \beta_1 + \frac{\beta_2 D_{\mathbf{u}}X_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} \\ &= \beta_1 + \beta_2\alpha_1 + \frac{\beta_2 D_{\mathbf{u}}\tilde{w}_2(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}\tilde{w}_1(\mathbf{s})} \end{aligned} \tag{4.7}$$

$$\begin{aligned} \frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} &= \beta_2 + \frac{\beta_1 D_{\mathbf{u}}X_1(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} \\ &= \beta_2 + \beta_1\frac{D_{\mathbf{u}}\tilde{w}_1(\mathbf{s})}{\alpha_1 D_{\mathbf{u}}\tilde{w}_1(\mathbf{s}) + D_{\mathbf{u}}\tilde{w}_2(\mathbf{s})} + \frac{D_{\mathbf{u}}w_y(\mathbf{s})}{\alpha_1 D_{\mathbf{u}}\tilde{w}_1(\mathbf{s}) + D_{\mathbf{u}}\tilde{w}_2(\mathbf{s})} \end{aligned} \tag{4.8}$$

The ratio in (4.7) is a ratio of mean-zero independent Normal random variables,

similar to the sensitivity for the bivariate model. The sensitivity in (4.8) is written as β_2 plus a sum of two dependent ratios; the first ratio has dependence between numerator and denominator, while the second is a ratio of two independent mean-zero normal random variables.

Similar to the unconditional specification, under some conditions these sensitivities are centered at the global coefficients for the variables. The global coefficients $\tilde{\beta}_{1|x_1}$ and $\tilde{\beta}_{2|x_2}$ are defined as

$$\begin{aligned} E(Y(\mathbf{s})|X_1(\mathbf{s})) &= (\beta_0 + \beta_2\alpha_0) + (\beta_1 + \beta_2\alpha_1)X_1(\mathbf{s}) \\ &= \tilde{\beta}_{0|x_1} + \tilde{\beta}_{1|x_1}X_1(\mathbf{s}) \end{aligned} \tag{4.9}$$

$$\begin{aligned} E(Y(\mathbf{s})|X_2(\mathbf{s})) &= (\eta_0 - \frac{\alpha_1\sigma_1^2(\alpha_0 + \alpha_1\eta_0)}{\alpha_1^2\sigma_1^2 + \sigma_2^2}) + (\beta_2 + \beta_1\frac{\alpha_1}{\alpha_1^2 + u_2^2/u_1^2})X_2(\mathbf{s}) \\ &= \tilde{\beta}_{0|x_2} + \tilde{\beta}_{2|x_2}X_2(\mathbf{s}) \end{aligned} \tag{4.10}$$

where $Var(\tilde{w}_i(\mathbf{s})) = u_i^2$.

It is immediately clear that the centering for (4.7) is equal to the global coefficient defined in (4.9), and this centering does not depend on the variance parameters for the processes. The centering for (4.8) as the global coefficient can be seen by showing that the ratio $D_{\mathbf{u}}\tilde{w}_1(\mathbf{s})/(\alpha_1 D_{\mathbf{u}}\tilde{w}_1(\mathbf{s}) + D_{\mathbf{u}}\tilde{w}_2(\mathbf{s}))$ is centered at $\alpha_1/(\alpha_1^2 + \tilde{v}_2^2/\tilde{v}_1^2)$ where $Var(D_{\mathbf{u}}\tilde{w}_i(\mathbf{s})) = \tilde{v}_i^2$. Thus, the centering will be equal to the global coefficient if $\tilde{v}_2^2/\tilde{v}_1^2 = u_2^2/u_1^2$.

For a Matérn covariance $Var(D_{\mathbf{u}}w_i(\mathbf{s})) = u_i^2\phi_i^2$, and equivalence will occur when the covariates share a common range parameter $\phi_1 = \phi_2$. This is equivalent to the result for the unconditional model, but here applies only to one of the two covariates.

When contemplating fitting a multivariate model under the conditional or unconditional specification it is important to consider whether the variables can reasonably be assumed to be ordered. The unconditional specification produces symmetric inference for the two variables, with sensitivities centered at the global coefficient

contingent on the process variances. In contrast, under the conditional specification the sensitivity for the variable modeled independently, $X_1(\mathbf{s})$ here, will always be centered at the global relationship parameter regardless of the process variances, while the centering for the other sensitivity will be contingent on the process variances. This asymmetry between the variables may be undesirable when no clear ordering exists.

For either specification of the model one can additionally compute the spatial angular discrepancy process associated with each of the covariates,

$$disc_1(\mathbf{s}) = 1 - cos(\theta_1(\mathbf{s}) - \theta_y(\mathbf{s}))$$

$$disc_2(\mathbf{s}) = 1 - cos(\theta_2(\mathbf{s}) - \theta_y(\mathbf{s}))$$

where $\theta_i(\mathbf{s})$ is the angle describing the direction of maximum gradient for the $X_i(\mathbf{s})$ or $Y(\mathbf{s})$ surface at location $\mathbf{s}$. These processes inform on the strength of the relationship between variables, with a value of 0 indicating a strong positive relationship and a value of 2 indicating a strong negative relationship. Because the directions of maximum gradient are computed based on the full joint model, the interpretations regarding the response-covariate relationships will be conditional on the additional variables in the model.

4.2.2 Multiple Responses

Consider a pair of correlated response surfaces $\{Y_1(\mathbf{s}), Y_2(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ that are both dependent on a single covariate surface $\{X(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$. Let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))'$ and $\mathbf{Y}(\mathbf{s}) = \boldsymbol{\gamma} + X(\mathbf{s})\boldsymbol{\beta} + \mathbf{A}\mathbf{W}(\mathbf{s})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$, $\mathbf{A}$ is a symmetric 2×2 matrix and $\mathbf{W}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}))'$ is a pair of independent zero-mean Gaussian processes with spatial covariance functions $G_1(\boldsymbol{\delta}; 1, \phi_1)$ and $G_2(\boldsymbol{\delta}; 1, \phi_2)$ respectively. The covariate surface is modeled as a spatially varying intercept using a Gaussian process with covariance function $K(\boldsymbol{\delta}; \sigma_y^2, \phi_y)$ with some mean $E(X(\mathbf{s})) = \eta_0$. Observed at a set of locations $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1), \ldots, \mathbf{Y}(\mathbf{s}_n))$, we write $\mathbf{Y}|\mathbf{X} \sim N(\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta}, \mathbf{G_1} \otimes \mathbf{A_1} +$

$\mathbf{G_2} \otimes \mathbf{A_2}$) and $\mathbf{X} \sim N(\eta_0, \mathbf{K})$. Similar to the multi-covariate version discussed in 4.2.1, this model can be extended to accommodate additional response variables through the introduction of an additional zero-mean Gaussian process $w_i(\mathbf{s})$ for each additional variable. The general model for $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}), \ldots, Y_m(\mathbf{s}))'$ would then be $\mathbf{Y} \sim N(\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta}, \Sigma_{i=1}^{m} \mathbf{G_i} \otimes \mathbf{A_i})$. Prior distributions for the parameters would be comparable to those discussed in 4.2.1.

To facilitate computation of the directional derivatives $D_{\mathbf{u}}Y_1(\mathbf{s}), D_{\mathbf{u}}Y_2(\mathbf{s})$ and $D_{\mathbf{u}}X(\mathbf{s})$ the model is rewritten in terms of independent zero-mean Gaussian processes,

$$\begin{aligned} Y_1(\mathbf{s}) &= \gamma_1 + \beta_1 X(\mathbf{s}) + a_{11} w_1(\mathbf{s}) + a_{12} w_2(\mathbf{s}) \\ Y_2(\mathbf{s}) &= \gamma_2 + \beta_2 X(\mathbf{s}) + a_{12} w_1(\mathbf{s}) + a_{22} w_2(\mathbf{s}) \\ X(\mathbf{s}) &= \eta_0 + w_x(\mathbf{s}) \end{aligned} \tag{4.11}$$

where $w_1(\mathbf{s})$, $w_2(\mathbf{s})$ and $w_x(\mathbf{s})$ are independent Gaussian processes. This representation of the model allows us to easily compute the associated local directional sensitivity processes,

$$\frac{D_{\mathbf{u}}Y_1(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \beta_1 + \frac{a_{11} D_{\mathbf{u}} w_1(\mathbf{s}) + a_{12} D_{\mathbf{u}} w_2(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \tag{4.12}$$

$$\frac{D_{\mathbf{u}}Y_2(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \beta_2 + \frac{a_{12} D_{\mathbf{u}} w_1(\mathbf{s}) + a_{22} D_{\mathbf{u}} w_2(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \tag{4.13}$$

Each of these sensitivities is a ratio of independent normal random variables and follows a Cauchy distribution centered at the parameter defining the global relationship between response $Y_i(\mathbf{s})$ and the covariate $X(\mathbf{s})$. In this case the global coefficient is simply the coefficient of $X(\mathbf{s})$ in the spatial linear model. This result is equivalent to the result for the bivariate model.

The conditional specification for multiple-response data does not suffer from the same asymmetry downfall as it does for the multiple-covariate data. However, the

dependence structure is inherited by the local directional sensitivities in a manner that, again, may be undesirable when no inherent ordering of the responses exists. The model can be written,

$$\begin{aligned} Y_1(\mathbf{s}) &= \tilde{\beta}_0 + \tilde{\beta}_1 Y_2(\mathbf{s}) + \tilde{\beta}_2 X(\mathbf{s}) + \tilde{w}_1(\mathbf{s}) \\ Y_2(\mathbf{s}) &= \tilde{\alpha}_0 + \tilde{\alpha}_1 X(\mathbf{s}) + \tilde{w}_2(\mathbf{s}) \\ X(\mathbf{s}) &= \eta_0 + w_x(\mathbf{s}) \end{aligned} \tag{4.14}$$

where $\tilde{w}_1(\mathbf{s}), \tilde{w}_2(\mathbf{s})$ and $w_x(\mathbf{s})$ are mean-zero independen Gaussian processes. The resulting sensitivities are

$$\frac{D_{\mathbf{u}} Y_2(\mathbf{s})}{D_{\mathbf{u}} X(\mathbf{s})} = \tilde{\alpha}_1 + \frac{D_{\mathbf{u}} \tilde{w}_2(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \tag{4.15}$$

$$\begin{aligned} \frac{D_{\mathbf{u}} Y_1(\mathbf{s})}{D_{\mathbf{u}} X(\mathbf{s})} &= \tilde{\beta}_2 + \tilde{\beta}_1 \tilde{\alpha}_1 + \frac{\tilde{\beta}_1 D_{\mathbf{u}} \tilde{w}_2(\mathbf{s}) + D_{\mathbf{u}} \tilde{w}_1(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \\ &= \tilde{\beta}_2 + \tilde{\beta}_1 \frac{D_{\mathbf{u}} Y_2(\mathbf{s})}{D_{\mathbf{u}} X(\mathbf{s})} + \frac{D_{\mathbf{u}} \tilde{w}_1(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} \end{aligned} \tag{4.16}$$

The above sensitivities are clearly centered at the parameters defining the global relationships between response-covariate pairs, $\tilde{\alpha}_1$ and $\tilde{\beta}_{1|x}$, where

$$\begin{aligned} E(Y_1(\mathbf{s})|X(\mathbf{s})) &= (\tilde{\beta}_0 + \tilde{\beta}_1 \tilde{\alpha}_0) + (\tilde{\beta}_2 + \tilde{\beta}_1 \tilde{\alpha}_1) X(\mathbf{s}) \\ &= \tilde{\beta}_{0|x} + \tilde{\beta}_{1|x} X(\mathbf{s}) \end{aligned} \tag{4.17}$$

While there is no asymmetry in the centering of the sensitivities, the dependence between the responses is apparent in the form of the sensitivity processes as shown in (4.16). For the $Y_1(\mathbf{s})$ variable the local relationship with the covariate is dependent on the spatial relationship of $Y_1(\mathbf{s})$ and $X(\mathbf{s})$, but it is also dependent on the spatial relationship between the covariate with the other response variable $Y_2(\mathbf{s})$. A symmetric pattern does not hold true when examining the local relationship between $Y_2(\mathbf{s})$ and the covariate, as shown in (4.15); the local relationship between $X(\mathbf{s})$ and $Y_2(\mathbf{s})$ depends only on the spatial structure of those two variables. This asymmetry

in the treatment of the response variables, and its desirability, should be taken into account when specifying a multivariate model for multiple responses.

Once again, the directional derivatives can additionally be used to compute the spatial angular discrepancy process associated with each of the responses and will exhibit similar properties. We have

$$disc_1(\mathbf{s}) = 1 - cos(\theta_x(\mathbf{s}) - \theta_1(\mathbf{s}))$$

$$disc_2(\mathbf{s}) = 1 - cos(\theta_x(\mathbf{s}) - \theta_2(\mathbf{s}))$$

where $\theta_i(\mathbf{s})$ is the angle describing the direction of maximum gradient for the $Y_i(\mathbf{s})$ or $X(\mathbf{s})$ surface at any location $\mathbf{s}$. The directions of maximum gradient are computed based on surfaces from the joint model, so resulting inference will be conditional on the other model variables.

4.2.3 Multiple Responses and Predictors

Consider a pair of correlated response surfaces $\{Y_1(\mathbf{s}), Y_2(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ that are dependent on a pair of correlated covariate surfaces $\{X_1(\mathbf{s}), X_2(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$. Let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))'$ and $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), X_2(\mathbf{s}))'$. Define the dependence between response and covariate such that $\mathbf{Y}(\mathbf{s}) = \boldsymbol{\gamma} + \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \mathbf{AW}(\mathbf{s})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$, $\mathbf{A}$ is a symmetric 2×2 matrix and $\mathbf{W}(\mathbf{s}) = (w_1(\mathbf{s}), w_2(\mathbf{s}))'$ is a pair of independent zero-mean Gaussian processes with spatial covariance functions $K_1(\boldsymbol{\delta}; 1, \phi_{y1})$ and $K_2(\boldsymbol{\delta}; 1, \phi_{y2})$ respectively. The covariate surfaces are modeled similarly such that $\mathbf{X}(\mathbf{s}) = \boldsymbol{\eta} + \mathbf{BV}(\mathbf{s})$, where $\boldsymbol{\eta} = (\eta_1, \eta_2)'$, $\mathbf{B}$ is a symmetric 2×2 matrix and $\mathbf{V}(\mathbf{s}) = (v_1(\mathbf{s}), v_2(\mathbf{s}))'$ is a pair of independent zero-mean Gaussian processes with spatial covariance functions $G_1(\boldsymbol{\delta}; 1, \phi_{x1})$ and $G_2(\boldsymbol{\delta}; 1, \phi_{x2})$. Observed at a set of locations $\mathbf{Y} = (\mathbf{Y}(s_1), \ldots, \mathbf{Y}(s_n))$, we write $\mathbf{Y}|\mathbf{X} \sim N(\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta}, \mathbf{K_1} \otimes \mathbf{A_1} + \mathbf{K_2} \otimes \mathbf{A_2})$ and $\mathbf{X} \sim N(\boldsymbol{\eta}, \mathbf{G_1} \otimes \mathbf{B_1} + \mathbf{G_2} \otimes \mathbf{B_2})$. Additional response and covariate variables can be incorporated as discussed in Sections 4.2.1 and 4.2.2.

Directional derivatives $D_{\mathbf{u}}Y_1(\mathbf{s}), D_{\mathbf{u}}Y_2(\mathbf{s}), D_{\mathbf{u}}X_1(\mathbf{s})$ and $D_{\mathbf{u}}X_2(\mathbf{s})$ can be calculated using samples of the gradient processes for the independent zero-mean Gaussian processes used to define the model,

$$\begin{aligned} Y_1(\mathbf{s}) &= \gamma_1 + \beta_{11}X_1(\mathbf{s}) + \beta_{12}X_2(\mathbf{s}) + a_{11}w_1(\mathbf{s}) + a_{12}w_2(\mathbf{s}) \\ Y_2(\mathbf{s}) &= \gamma_2 + \beta_{21}X_1(\mathbf{s}) + \beta_{22}X_2(\mathbf{s}) + a_{12}w_1(\mathbf{s}) + a_{22}w_2(\mathbf{s}) \\ X_1(\mathbf{s}) &= \eta_1 + b_{11}v_1(\mathbf{s}) + b_{12}v_2(\mathbf{s}) \\ X_2(\mathbf{s}) &= \eta_2 + b_{12}v_1(\mathbf{s}) + b_{22}v_2(\mathbf{s}) \end{aligned} \tag{4.18}$$

where $w_1(\mathbf{s})$, $w_2(\mathbf{s})$, $v_1(\mathbf{s})$ and $v_2(\mathbf{s})$ are independent Gaussian processes. Expressions for computing the associated local directional sensitivities are similar to those in Section 4.2.1 and are provided here for the $Y_1(\mathbf{s})$ process,

$$\begin{aligned} \frac{D_{\mathbf{u}}Y_1(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} &= \beta_{11} + \frac{\beta_{12}D_{\mathbf{u}}X_2(\mathbf{s}) + a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{12}D_{\mathbf{u}}w_2(\mathbf{s})}{D_{\mathbf{u}}X_1(\mathbf{s})} \\ &= \beta_{11} + \frac{\beta_{12}b_{11}D_{\mathbf{u}}v_1(\mathbf{s}) + \beta_{12}b_{12}D_{\mathbf{u}}v_2(\mathbf{s}) + a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{12}D_{\mathbf{u}}w_2(\mathbf{s})}{b_{11}D_{\mathbf{u}}v_1(\mathbf{s}) + b_{12}D_{\mathbf{u}}v_2(\mathbf{s})} \end{aligned} \tag{4.19}$$

$$\begin{aligned} \frac{D_{\mathbf{u}}Y_1(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} &= \beta_{12} + \frac{\beta_{11}D_{\mathbf{u}}X_1(\mathbf{s}) + a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{12}D_{\mathbf{u}}w_2(\mathbf{s})}{D_{\mathbf{u}}X_2(\mathbf{s})} \\ &= \beta_{12} + \frac{\beta_{11}b_{12}D_{\mathbf{u}}v_1(\mathbf{s}) + \beta_{11}b_{22}D_{\mathbf{u}}v_2(\mathbf{s}) + a_{11}D_{\mathbf{u}}w_1(\mathbf{s}) + a_{12}D_{\mathbf{u}}w_2(\mathbf{s})}{b_{12}D_{\mathbf{u}}v_1(\mathbf{s}) + b_{22}D_{\mathbf{u}}v_2(\mathbf{s})} \end{aligned} \tag{4.20}$$

Sensitivities for the $Y_2(\mathbf{s})$ process are analogous.

Using arguments similar to those in Section 4.2.1 it can be shown that these ratios will be centered at the global coefficients relating each covariate to each response variable. The global coefficient relating $Y_1(\mathbf{s})$ and $X_1(\mathbf{s})$ is defined as $\beta^{y1}_{1|x1}$, where

$$\begin{aligned} E(Y_1(\mathbf{s})|X_1(\mathbf{s})) &= (\gamma_1 + \beta_{12}\eta_2 - \beta_{12}\frac{b_{11}b_{12} + b_{22}b_{12}}{b_{11}^2 + b_{12}^2}\eta_1) + (\beta_{11} + \frac{b_{11}b_{12} + b_{22}b_{12}}{b_{11}^2 + b_{12}^2})X_1(\mathbf{s}) \\ &= \beta^{y1}_{0|x1} + \beta^{y1}_{1|x1}X_1(\mathbf{s}) \end{aligned}$$

Global coefficients relating the other responses and covariates can be computed in the same manner.

There will be a spatial angular discrepancy process associated with each pair of response and covariate,

$$disc_{11}(\mathbf{s}) = 1 - cos(\theta_{x1}(\mathbf{s}) - \theta_{y1}(\mathbf{s}))$$

$$disc_{12}(\mathbf{s}) = 1 - cos(\theta_{x1}(\mathbf{s}) - \theta_{y2}(\mathbf{s}))$$

$$disc_{21}(\mathbf{s}) = 1 - cos(\theta_{x2}(\mathbf{s}) - \theta_{y1}(\mathbf{s}))$$

$$disc_{22}(\mathbf{s}) = 1 - cos(\theta_{x2}(\mathbf{s}) - \theta_{y2}(\mathbf{s}))$$

where $\theta_{xi}(\mathbf{s})$ and $\theta_{yi}(\mathbf{s})$ are the angles describing the direction of maximum gradient for the $Y_i(\mathbf{s})$ or $X_i(\mathbf{s})$ surface at location $\mathbf{s}$. The spatial angular discrepancies hold similar properties as described earlier, and will again provide inference conditional on the variables included in the joint model.

4.3 Multivariate Examples using South African Plant Traits

4.3.1 Data

A dataset coming from the Cape Floristic Region in South Africa is used to illustrate the gradient analyses outlined in Sections 4.2.1 and 4.2.2 for models with multiple covariates or multiple responses of interest. The data was collected for 37 plant species in the genus *Protea* in order to assess plant trait patterns associated with environmental covariates in the region. Data are available from 36 plots, with inter-plot distances ranging from 78 to 366,100 meters apart. Figure 4.1 provides the plot locations. The poor spatial spread of the plots prevents any meaningful inference on the shape of the spatial surfaces, but the multivariate nature of the data lends itself to illustrating the gradient methodology for multiple predictors or responses. Within each plot leaf trait measurements are available for one or more *Protea* species, with the total number of plants measured per plot ranging from 5 to 24. Many species are

only available in a single plot, and generally 8 individuals are sampled within each species.

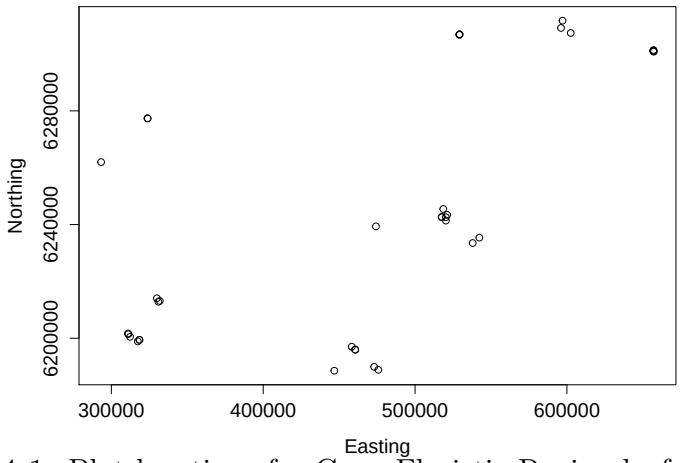

FIGURE 4.1: Plot locations for Cape Floristic Region leaf trait data.

As response variables we consider two plant traits, leaf dry mass per unit area (LMA) and leaf fresh water content (FWC). LMA (multiplied by 100 for convenience) ranges between 0.7 and 8.9 with plot-level averages between 1.3 and 4.6. FWC ranges between 0.5 and 3.7 with plot-level averages between 0.9 and 2.4. See the top row of Figure 4.2 for associated histograms. As covariates we consider two climate variables, mean annual temperature (MAT) and mean annual precipitation (MAP). MAT ranges between 11 and 18 degrees Celsius, and MAP ranges between 331 and 1332. See the bottom row of Figure 4.2 for associated histograms.

4.3.2 Multiple Predictors

To illustrate a multiple-covariate model let the response variable $Y(\mathbf{s})$ be leaf mass per area (LMA), and the covariates $X_1(\mathbf{s})$ and $X_2(\mathbf{s})$ be mean annual temperature (MAT) and mean annual precipitation (MAP). As described in Section 4.3.1, several measurements are available for the response variable at each location. Denote these observations as $Y_i(\mathbf{s})$ for $i = 1, \ldots, n_s$, where n_s is the number of observations available at location $\mathbf{s}$. These data are jointly modeled using a coregionalization

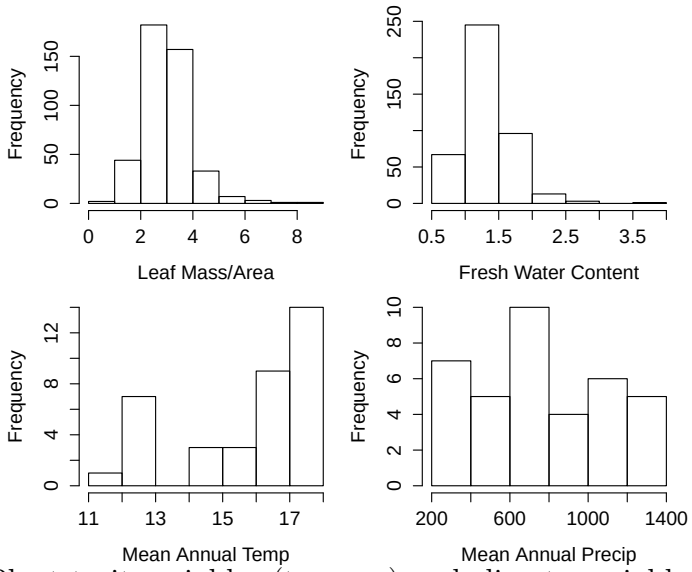

FIGURE 4.2: Plant trait variables (top row) and climate variables (bottom row).

specification for the covariates,

$$\begin{aligned} Y_i(\mathbf{s}) &= \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + w_y(\mathbf{s}) + \epsilon_i(\mathbf{s}) \\ &= \mu(\mathbf{s}) + \epsilon_i(\mathbf{s}) \\ X_1(\mathbf{s}) &= \gamma_1 + a_{11} w_1(\mathbf{s}) + a_{12} w_2(\mathbf{s}) \\ X_2(\mathbf{s}) &= \gamma_2 + a_{12} w_1(\mathbf{s}) + a_{22} w_2(\mathbf{s}) \end{aligned} \tag{4.21}$$

where $w_y(\mathbf{s})$, $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$ are independent Gaussian processes with Matérn covariance parameters (ϕ_y, σ_y^2), $(\phi_1, 1)$ and $(\phi_2, 1)$. The variance parameters σ^2 in the covariance function are set to 1 for the $X_1(\mathbf{s})$ and $X_2(\mathbf{s})$ processes to ensure identifiability of the coefficients.

This model is adapted from the one discussed in Section 4.2.1 to accommodate the repeated measurements at each location by introducing a non-spatial error term. For each plot the observed responses $Y_i(\mathbf{s})$ are assumed to share a common mean $\mu(\mathbf{s})$, with some individual deviation from this mean represented by independent $\epsilon_i(\mathbf{s}) \sim N(0, \tau^2)$. In the gradient analyses, the directional derivatives of interest will be associated with the mean surface $\mu(\mathbf{s})$ and and not with the individual $Y_i(\mathbf{s})$.

Otherwise, the local directional sensitivities will be as described in Section 4.2.1

4.3.3 Multiple Responses

To illustrate a multiple-response model consider $Y_1(\mathbf{s})$ as leaf mass per area (LMA) and $Y_2(\mathbf{s})$ as fresh water content (FWC), with the covariate mean annual temperature (MAT) as $X(\mathbf{s})$. As described above, the response variables have several measurements at each location, accommodated through the introduction of a non-spatial error term $\epsilon(\mathbf{s})$. The subsequent gradient analyses then consider the relationship between the mean responses $\mu_1(\mathbf{s})$ and $\mu_2(\mathbf{s})$ and the covariate. The model is written using a coregionalization specification for the responses,

$$\begin{aligned} Y_{1i}(\mathbf{s}) &= \beta_0 + \beta_1 X(\mathbf{s}) + a_{11} w_1(\mathbf{s}) + a_{12} w_2(\mathbf{s}) + \epsilon_{1i}(\mathbf{s}) \\ &= \mu_1(\mathbf{s}) + \epsilon_{1i}(\mathbf{s}) \\ Y_{2i}(\mathbf{s}) &= \alpha_0 + \alpha_1 X(\mathbf{s}) + a_{12} w_1(\mathbf{s}) + a_{22} w_2(\mathbf{s}) + \epsilon_{2i}(\mathbf{s}) \\ &= \mu_2(\mathbf{s}) + \epsilon_{2i}(\mathbf{s}) \\ X(\mathbf{s}) &= \eta_0 + w_x(\mathbf{s}) \end{aligned} \tag{4.22}$$

where $w_1(\mathbf{s})$, $w_2(\mathbf{s})$ and $w_x(\mathbf{s})$ are independent Gaussian processes with Matérn covariance parameters $(\phi_1, 1)$, $(\phi_2, 1)$, and (ϕ_x, σ_x^2). The variance parameters σ^2 for the $Y_i(\mathbf{s})$ processes are set to 1 for identifiability.

The local directional sensitivity processes are as described in Section 4.2.2, except that they are now associated with $D_{\mathbf{u}}\mu_1(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ and $D_{\mathbf{u}}\mu_2(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$.

4.4 Results

4.4.1 Multiple Predictors

Posterior means and 95% credible intervals for the model parameters in (4.21) are provided in Table 4.1. Estimates are based on 5000 samples, after a burn-in of 1000 iterations and a thinning of every 10th iterate. The credible intervals for both

covariates overlap zero, but MAT appears to have a more meaningful effect than MAP. The point estimate for MAP is effectively zero and the interval is nearly symmetric around zero, suggesting no effect on LMA. The posterior mean kriged surfaces for the three variables $X_1(\mathbf{s})$, $X_2(\mathbf{s})$ and $\mu(\mathbf{s})$ are provided in Figure 4.3 and exhibit limited spatial structure as a result of the spatial spread in the observed data.

Table 4.1: Parameter estimates and 95% credible intervals for the $[Y|X_1, X_2][X_1, X_2]$ model.

	Estimate	$q_{0.025}$	$q_{0.975}$
β_0	4.3206	1.4759	7.3492
β_1	-0.0840	-0.2793	0.1079
β_2	0.000007	-0.0012	0.0012
σ_y^2	0.3323	0.0534	0.9328
ϕ_y	0.0006	0.0004	0.0010
τ_y^2	4.7344	2.0621	9.0188
γ_1	16.0338	15.0630	16.9828
γ_2	750.67	635.32	867.65
a_{11}	2.0258	1.5109	2.7432
a_{12}	0.4181	-0.2952	1.1540
a_{22}	249.50	199.54	314.76
ϕ_1	0.0007	0.0005	0.0010
ϕ_2	0.0005	0.0004	0.0007

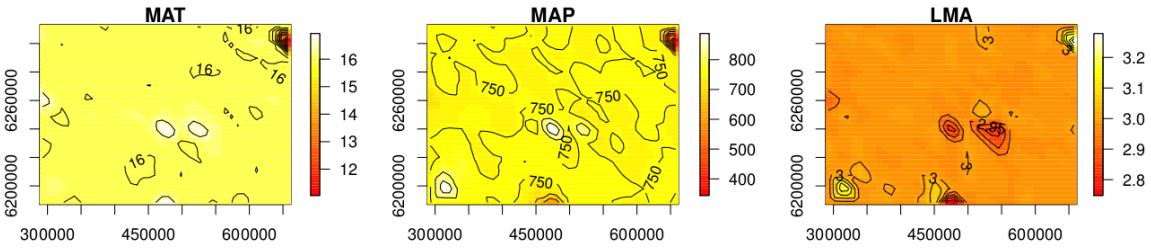

FIGURE 4.3: Posterior mean kriged surfaces for $X_1(\mathbf{s})$, $X_2(\mathbf{s})$, and $\mu(\mathbf{s})$ from left to right.

The posterior median local directional sensitivity surfaces and spatial angular discrepancy surfaces are provided in Figure 4.4. Unfortunately, the limited spatial spread in the observed data prevents illumination of any spatial story, but the processes enable general comparisons between the response variable and each of the

covariates. The local sensitivity corresponding to MAT, $D_{\mathbf{u}}\mu(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$, is negative everywhere, and the spatial discrepancy is roughly 1.5 everywhere. Both surfaces suggest that MAT has a negative effect on LMA, and in the case of the local sensitivity the effect corresponds to marginalizing over MAP. This marginal effect is not vastly different from the conditional effect suggested by $\beta_1 = -0.0840$, likely because the conditional effect of MAP is so close to zero. The local sensitivity for MAP, $D_{\mathbf{u}}\mu(\mathbf{s})/D_{\mathbf{u}}X_2(\mathbf{s})$, is negative almost everywhere and exhibits particularly strong negative values in the northeast corner. This suggests that marginally MAP has a negative effect on LMA. However, the spatial discrepancy is close to 1 everywhere, so after adjusting for MAT, any observed relationship between MAP and LMA is very weak.

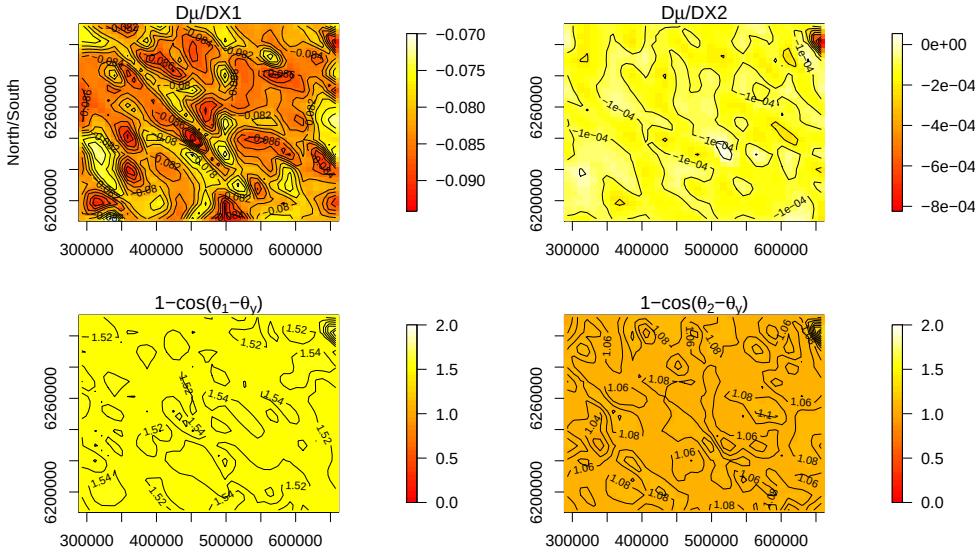

FIGURE 4.4: Top row: Posterior median local sensitivity surfaces; $D_{\mathbf{u}}\mu(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$ and $D_{\mathbf{u}}\mu(\mathbf{s})/D_{\mathbf{u}}X_2(\mathbf{s})$. Bottom row: Posterior median angular discrepancy surfaces; $disc_1(\mathbf{s})$ and $disc_2(\mathbf{s})$.

In both Figure 4.3 and Figure 4.4, the northeast corner stands out from the rest of the region. This corner corresponds to four plot locations in the observed data all coming from the Blesberg area of the Cape Floristic Region. The MAT and

MAP for the Blesberg plots are lower than for any other plots in the data, while the LMA measurements are close to average. The highlighting of this area in the local directional sensitivity for MAP indicates that the relationship between LMA and MAP is atypical in this region. This can most easily be seen through comparison with the southwest corner. The two areas have similar LMA, but correspond to opposite extremes for MAP. Since the data in the southwest corner aligns more closely with the rest of the data, the northwest corner is highlighted in the local directional sensitivity.

4.4.2 Multiple Responses

Posterior means and 95% credible intervals for the parameters in the model (4.22) are provided in Table 4.2. Estimates are based on 5000 samples after a burn-in of 1000 iterations and a thinning of every 10th iterate. The credible intervals for β_1 and α_1, relating each response variable to MAT, both overlap zero. However, the point estimates suggest MAT has a small negative effect on LMA and a small positive effect on MAP. The three posterior mean kriged surfaces for $\mu_1(\mathbf{s})$, $\mu_2(\mathbf{s})$ and $X_1(\mathbf{s})$ are provided in Figure 4.5, and again display limited spatial structure.

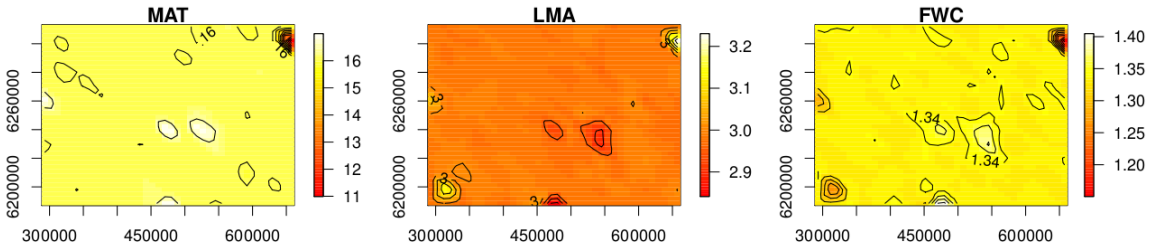

FIGURE 4.5: Posterior mean kriged surfaces for X, μ_1, and μ_2 from left to right.

The posterior median local directional sensitivities and spatial angular discrepancies are provided in Figure 4.6. As expected, the local sensitivity for LMA, $D_{\mathbf{u}}\mu_1(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$, is roughly centered at $\beta_1 = -0.0725$, the parameter defining the variables' relationship. The poor spatial structure of the observed data pre-

Table 4.2: Parameter estimates and 95% credible intervals for the $[Y_1, Y_2|X][X]$ model.

	Estimate	$q_{0.025}$	$q_{0.975}$
β_0	4.1415	1.1574	7.1401
β_1	-0.0725	-0.2629	0.1147
α_0	0.5275	-0.8393	1.7569
α_1	0.0502	-0.0251	0.1350
a_{11}	0.4567	0.0303	1.0044
a_{12}	-0.1501	-0.4203	0.0676
a_{22}	0.1812	0.0101	0.4064
τ_1^2	5.0906	2.1198	9.7001
τ_2^2	0.5453	0.1566	1.2413
ϕ_1	0.0006	0.0004	0.0009
ϕ_2	0.0006	0.0004	0.0010
η_0	16.0486	14.9481	17.0768
ϕ_x	0.0006	0.0004	0.0009
σ_x^2	5.3382	2.3916	11.5144

vents meaningful spatial variation beyond this centering. The angular discrepancy $disc_1(\mathbf{s})$ takes values close to 1.5, indicating a relatively strong negative relationship. The results for FWC are similar, but indicate a positive relationship. The local sensitivity $D_{\mathbf{u}}\mu_2(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ is roughly centered at $\alpha_1 = 0.0502$ with no meaningful spatial structure, and the angular discrepancy $disc_2(\mathbf{s})$ takes values of approximately 0.45 indicating a fairly strong positive relationship.

It's not surprising that the local directional sensitivity for LMA looks similar to the sensitivity $D_{\mathbf{u}}\mu(\mathbf{s})/D_{\mathbf{u}}X_1(\mathbf{s})$ in Figure 4.4 since both are describing the marginal effect of MAT on LMA. In this model MAT is the only covariate, while in the earlier model the local directional sensitivity was expected to be centered at the marginal effect of MAT after marginalizing over MAP. The similarity between these figures confirms the theoretical result.

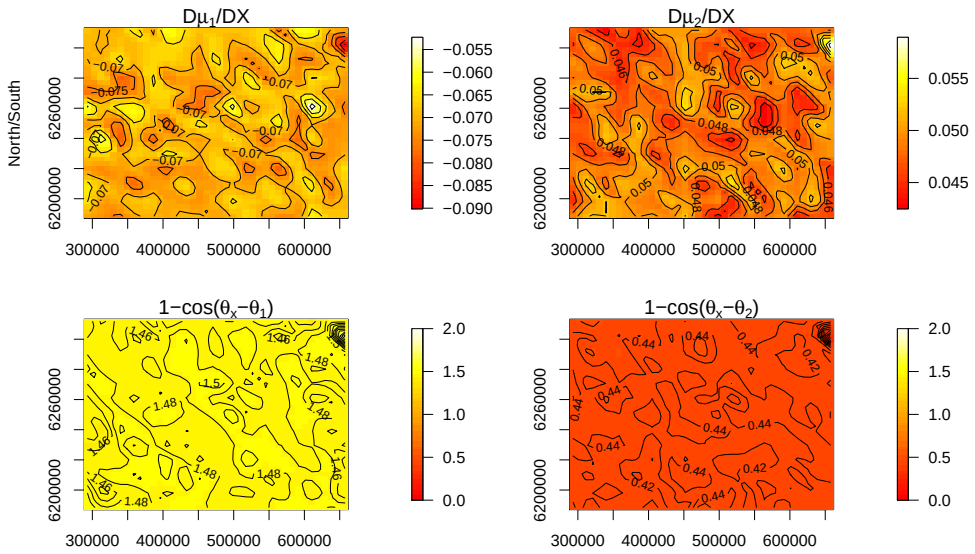

FIGURE 4.6: Top row: Posterior median local sensitivity surfaces; $D_{\mathbf{u}}\mu_1(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$ and $D_{\mathbf{u}}\mu_2(\mathbf{s})/D_{\mathbf{u}}X(\mathbf{s})$. Bottom row: Posterior median angular discrepancy surfaces; $disc_1(\mathbf{s})$, $disc_2(\mathbf{s})$.

4.5 Summary

The statistical methodology outlined in this chapter enables post-model fitting gradient analysis for multivariate spatial models. The local directional sensitivity process and spatial angular discrepancy process can be examined similarly to the bivariate framework, with slight adjustments in interpretation. When modeling a response using a single covariate, the local sensitivities will be centered at the parameter describing the global relationship between the variables. If additional covariates are introduced into the model the local sensitivity for any single covariate will be centered at the quantity defining the marginal effect of the covariate on the response. This marginal effect will generally *not* be equal to the fitted β parameter, since that parameter will define the variables' relationship conditioned on the other covariates in the model. The spatial angular discrepancy processes compare the directions of maximum gradient between any pair of surfaces, as before.

Inferentially, the joint multivariate model provides a global description of the

response-covariate pairwise relationships, conditional on the remaining set of variables. The post model-fitting gradient analyses expand on this description and compare the spatial variation in these relationships across variables. Similar to the model parameters, the spatial angular discrepancies inform on the sign and strength of pairwise behaviors conditional on the other covariates, while the local directional sensitivities inform on the marginal relationships between response and covariate. Together, these quantities provide a fuller picture of the dependencies suggested by the joint model.

The development in Section 4.2.1 highlights the importance of model choice for coregionalization models. Banerjee et al. (2004) showed that the unconditional (symmetric) and conditional (upper triangular) specifications can be considered reparameterizations of the same model, but inference regarding rates of change will vary under the two models. The unconditional model treats the variables identically, each dependent on the other, and this is apparent in the resulting directional derivatives. Under the conditional model, one variable is modeled independently from the other and again this shows in the resulting directional derivatives. When a clear ordering exists between the variables, a conditional model will be a reasonable choice, but if no obvious ordering can be assumed then the unconditional model will provide the desired symmetric treatment of the variables.

The methodology was demonstrated for plant trait and climate data coming from the Cape Floristic Region in South Africa, illustrating both a multiple-response model and a multiple-covariate model. The presence of several correlated response variables (plant traits) and several correlated covariates (climate variables) in these data necessitated the use of multivariate spatial models. The plant trait measurements cannot reasonably be assumed independent from one another, nor can one sensibly assign an order of dependence to the traits, so a coregionalization model with a symmetric specification was deemed most appropriate for their joint model.

A similar reasoning regarding temperature and precipitation encouraged a similar model be used for their joint model.

A few benefits of the gradient analysis were highlighted in the multiple-covariate model, where leaf mass per area is assumed to be a function of precipitation and temperature. Although the fitted model suggests that there is no real relationship with precipitation after adjusting for temperature, the local directional sensitivity suggests that marginally precipitation would have a negative effect. This negative effect is in line with results described by Lamont et al. (2002). The fitted model also suggested a negative effect for temperature, both marginally and conditional on precipitation. This is also in line with current literature, as described in Poorter et al. (2009). The local sensitivities additionally highlighted the Blesberg region as potentially behaving differently from the other locations, a difference that could be further investigated if deemed to be of interest.

5

Additional Gradient Analysis Extensions

5.1 Spatial-Temporal Modeling

5.1.1 Introduction

In addition to spatial structure, many researchers will be interested in learning how variable relationships evolve temporally. A readily available environmental example would be measurements of ozone and temperature, which are often available on a daily time scale. Accommodating such data requires flexible models allowing for both spatial and temporal dynamics, which have a rich presence in the literature (e.g. Wikle et al., 1998; Banerjee et al., 2004; Gelfand et al., 2005; Cressie and Wikle, 2011; Reich et al., 2011b). So called "separable" models can be factored into a purely spatial component and a purely temporal component. This is generally undesirable because it does not allow for space-time interactions that would be expected for many data. The class of "non-separable" models are more complicated to fit but allow for more complex spatial-temporal dynamics to be captured since they cannot be factored in the same way. In addition, models can be classified based on whether they discretize space and/or time, with continuous space and discrete time perhaps

most common.

Process gradients were recently explored for spatial-temporal models by Quick et al. (2013) to identify time points with significantly high residual gradients that would suggest a missing covariate. The scenario they consider involves temporal gradients in the continuous time and discrete space setting. In contrast, the methodology we propose examines spatial gradients in the discrete time and continuous space setting. Their methodology assesses temporal gradients after adjusting for spatial effects, while ours assesses spatial gradients after adjusting for temporal effects. Either approach may be desirable depending on the setting.

To illustrate spatial gradient analyses for spatial-temporal models we consider three fairly basic forms for a spatial-temporal error term with increasing complexity. In addition to the relative rate of change between response-covariate pairs, the temporal setting enables comparisons of the relative rates of change for a single surface at multiple time points. This time-lag comparison provides insight into the temporal evolution for the surface's rate of change.

5.1.2 Modeling and Gradient Theory

Consider a response $Y(\mathbf{s}, t)$ and covariate $X(\mathbf{s}, t)$ each observed at a set of locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and a set of time points $t_1, \ldots, t_m$. Assume the following joint spatial-temporal model,

$$\begin{aligned} Y(\mathbf{s}, t) &= \beta_0 + \beta_1 X(\mathbf{s}, t) + e_y(\mathbf{s}, t) \\ X(\mathbf{s}, t) &= \alpha_0 + e_x(\mathbf{s}, t) \end{aligned} \tag{5.1}$$

The error terms for the $Y(\mathbf{s}, t)$ and $X(\mathbf{s}, t)$ processes capture the spatial and temporal dependence in one of three ways:

$$e(\mathbf{s}, t) = \eta(t) + w(\mathbf{s}) \tag{5.2}$$

$$e(\mathbf{s}, t) = w^{(t)}(\mathbf{s}) \tag{5.3}$$

$$e(\mathbf{s}, t) = v(\mathbf{s}, t) \tag{5.4}$$

The error structure in (5.2) assumes independent spatial and temporal effects where $w(\mathbf{s}) \sim GP(0, \rho(\cdot))$ is shared across all time points and $\eta(t) = \eta(t-1) + \epsilon(t)$ with *iid* $\epsilon(t) \sim N(0, \tau^2)$. Next, the error structure in (5.3) assumes a unique independent spatial effect at each time point with $w^{(t)}(\mathbf{s}) \sim GP(0, \rho(\cdot))$. Finally, the error structure in (5.4) allows for a spatial and temporal evolution at each time point such that $v(\mathbf{s}, t) = v(\mathbf{s}, t-1) + w^{(t)}(\mathbf{s}) + \epsilon(t)$ with $w^{(t)}(\mathbf{s})$ and $\epsilon(t)$ defined as in the previous models.

Note, a pure (non-spatial and non-temporal) error component could be added to any of the above error structures, but for the gradient analyses would be ignored with directional derivatives being calculated on the remaining smooth surface.

Spatial gradients can be computed for each of the models described above and not surprisingly inherit the characteristics present in the parent model. Similar to the non-temporal models discussed in previous chapters, these can then be used to compute the local directional sensitivities comparing the the spatial rates of change at any given time point. In addition, local directional time-lag ratios can be considered for each variable, and may give insight into how the surface's rates of change vary over time.

The local directional sensitivities at time t under each of the three error structures

will be

$$\frac{D_{\mathbf{u}}Y(\mathbf{s},t)}{D_{\mathbf{u}}X(\mathbf{s},t)} = \beta_1 + \frac{D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})}$$

$$\frac{D_{\mathbf{u}}Y(\mathbf{s},t)}{D_{\mathbf{u}}X(\mathbf{s},t)} = \beta_1 + \frac{D_{\mathbf{u}}w_y^{(t)}(\mathbf{s})}{D_{\mathbf{u}}w_x^{(t)}(\mathbf{s})}$$

$$\frac{D_{\mathbf{u}}Y(\mathbf{s},t)}{D_{\mathbf{u}}X(\mathbf{s},t)} = \beta_1 + \frac{D_{\mathbf{u}}v_y(\mathbf{s},t)}{D_{\mathbf{u}}v_x(\mathbf{s},t)}$$

All three local directional sensitivities will be Cauchy distributed centered at the global relationship parameter β_1. The Cauchy distribution can be seen by observing that in each case the ratio term in the sensitivity will involve two mean-zero independent normal random variables. The local directional sensitivity for (5.2) is constant through time, reflecting the time-invariant spatial error. Similarly, the sensitivity for (5.3) varies at each time point, but is independent across time. For (5.4), the autoregressive nature of the error function is inherited by the local directional sensitivity process. Substituting terms according to the definition of $v(\mathbf{s},t)$ we have

$$\begin{aligned}\frac{D_{\mathbf{u}}Y(\mathbf{s},t)}{D_{\mathbf{u}}X(\mathbf{s},t)} =& \beta_1 + \frac{D_{\mathbf{u}}v_y(\mathbf{s},t-1)}{D_{\mathbf{u}}v_x(\mathbf{s},t-1)} \\ &+ \frac{D_{\mathbf{u}}v_x(\mathbf{s},t-1)D_{\mathbf{u}}w_y^{(t)}(s) - D_{\mathbf{u}}v_y(\mathbf{s},t-1)D_{\mathbf{u}}w_x^{(t)}(s)}{D_{\mathbf{u}}v_x(\mathbf{s},t-1)(D_{\mathbf{u}}v_x(\mathbf{s},t-1) + D_{\mathbf{u}}w_x^{(t)}(s))} \\ =& \frac{D_{\mathbf{u}}Y(\mathbf{s},t-1)}{D_{\mathbf{u}}X(\mathbf{s},t-1)} + \frac{D_{\mathbf{u}}v_x(\mathbf{s},t-1)D_{\mathbf{u}}w_y^{(t)}(s) - D_{\mathbf{u}}v_y(\mathbf{s},t-1)D_{\mathbf{u}}w_x^{(t)}(s)}{D_{\mathbf{u}}v_x(\mathbf{s},t-1)(D_{\mathbf{u}}v_x(\mathbf{s},t-1) + D_{\mathbf{u}}w_x^{(t)}(s))}\end{aligned}$$

with the error term centered at zero.

The local directional time-lag ratios are uninteresting for (5.2) and (5.3) since neither model allows the spatial errors to evolve in time. Focusing on the third error structure, (5.4), the local directional time-lag ratio can be computed in two scenarios. First, when the mean is non-dynamic (i.e., $X(\mathbf{s},t)$ is not jointly modeled)

and interest is only in the behavior of the residual,

$$\frac{D_{\mathbf{u}}e_t(s)}{D_{\mathbf{u}}e_{t-1}(s)} = \frac{D_{\mathbf{u}}v(\mathbf{s},t)}{D_{\mathbf{u}}v(\mathbf{s},t-1)}$$
$$= 1 + \frac{D_{\mathbf{u}}w^{(t)}(\mathbf{s})}{D_{\mathbf{u}}v(\mathbf{s},t-1)}$$

and second, when the covariate $X(\mathbf{s},t)$ is jointly modeled,

$$\frac{D_{\mathbf{u}}Y_t(s)}{D_{\mathbf{u}}Y_{t-1}(s)} = \frac{\beta_1 D_{\mathbf{u}}X(\mathbf{s},t) + D_{\mathbf{u}}v_y(\mathbf{s},t)}{\beta_1 D_{\mathbf{u}}X(\mathbf{s},t-1) + D_{\mathbf{u}}v_y(\mathbf{s},t-1)}$$
$$= \frac{D_{\mathbf{u}}X_t(s)}{D_{\mathbf{u}}X_{t-1}(s)} + \frac{\beta_1 D_{\mathbf{u}}X(\mathbf{s},t-1)D_{\mathbf{u}}v_y(\mathbf{s},t) - \beta_1 D_{\mathbf{u}}X(\mathbf{s},t)D_{\mathbf{u}}v_y(\mathbf{s},t-1)}{\beta_1 D_{\mathbf{u}}X(\mathbf{s},t-1)(\beta_1 D_{\mathbf{u}}X(\mathbf{s},t-1) + D_{\mathbf{u}}v_y(\mathbf{s},t-1))}$$

In both scenarios the local directional time-lag ratio will be centered at 1, with some zero-centered error term. This is consistent with the belief that the rate of change for the spatial surface should roughly stay constant from year to year.

The examples provided here give some insight into the kinds of spatial gradients that can be examined in a spatial-temporal setting. Characteristics from the parent model will generally be inherited by the gradient processes, and the complexity of the gradient functions will depend on the complexity of the original model. We propose examination of both local directional sensitivities at a given time point, and local directional time-lag ratios across time points to gain a deeper understanding of the processes involved.

5.2 Data with Spatial Misalignment

5.2.1 Introduction

When collecting data for environmental or ecological studies there will often be spatial misalignment between measurements for climate variables and measurements for plant or animal characteristics. For example, observations for bud-burst will occur at the location of individual trees while temperature observations may be taken from

the nearest weather monitoring station or may correspond to the average temperature in a relevant spatial unit. Misalignment in the covariate and response observations, or between the observed spatial scale and desired inferential spatial scale, can inject errors and biases into the analysis. Several approaches have been proposed in the literature to accommodate such data (e.g. Mugglin et al., 2000; Gotway and Young, 2002; Banerjee et al., 2004; Gryparis et al., 2009; Peng and Bell, 2010).

Following discussions by Banerjee et al. (2004), we accommodate areal data by treating them as block averages across the region. This interpretation is appropriate for several potential climate measurements such as average temperature or precipitation, as well as potential environmental responses such as pollution. Areal data clearly do not constitute a smooth surface for which rates of change can be considered, a requirement for the gradient analyses presented thus far. Instead, we use the areal observations to make predictions at new point-referenced locations based on a Gaussian process model and the block average interpretation of the areal observations. This in turn allows for consideration of the rate of change for the corresponding surface, including comparisons with the rate of change for the response surface.

Here we illustrate how gradient analyses can be applied under three misalignment scenarios. The first is most straight forward and involves two point-referenced variables, response and covariate, that do not share a common set of observation locations. The second applies to scenarios when the covariate is only available as areal data, interpreted as a block average, and the response is point-referenced. Scenarios where the response is block averaged and the covariate is point-referenced would follow similarly. The final scenario assumes both the response and covariate are block average data.

5.2.2 Modeling and Gradient Theory

The methodology and notation for the misalignment scenarios follow those proposed by Banerjee et al. (2004).

For the first scenario we assume that the covariate and response variables were observed at different locations, e.g. tree locations vs. monitoring station locations. To understand the variables' relationship we would ideally have paired measurements, allowing the response to be regressed on corresponding covariate values at the same location. Instead, we observe the response $Y(\mathbf{s})$ at a set of locations $\mathcal{S}_y = \mathbf{s}_1, \ldots, \mathbf{s}_n$ and the covariate at a non-identical set of locations $\mathcal{S}_x = \mathbf{s}'_1, \ldots, \mathbf{s}'_m$. Consider a spatially varying intercept model

$$X(\mathbf{s}') = \alpha + w_x(\mathbf{s}')$$
$$w_x(\mathbf{s}') \sim GP(0, \rho_x(\sigma_x^2, \phi_x))$$

Letting $X_s = \{X(\mathbf{s}) : \mathbf{s} \in \mathcal{S}_y\}$ and $X_{s'} = \{X(\mathbf{s}') : \mathbf{s}' \in \mathcal{S}_x\}$, we can sample covariate values at the desired locations (i.e., the locations where $Y(\mathbf{s})$ was observed) using the predictive distribution

$$f(X_s|X_{s'}) = \int f(X_s|X_{s'}, \alpha, \sigma_x^2, \phi_x) f(\alpha_0, \sigma_x^2, \phi_x|X_{s'}) d\alpha d\sigma_x^2 d\phi_x$$

The spatial regression can be fit such that

$$Y(\mathbf{s}) = \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s})$$
$$w_y(\mathbf{s}) \sim GP(0, \rho_y(\sigma_y^2, \phi_y))$$

where $Y(\mathbf{s})$ and $X(\mathbf{s})$ are both mean square differentiable processes, and the model likelihood is

$$[Y(\mathbf{s})|X(\mathbf{s}), \boldsymbol{\beta}, \sigma_y^2, \phi_y][X(\mathbf{s})|X(\mathbf{s}'), \alpha, \sigma_x^2, \phi_x][X(\mathbf{s}')|\alpha, \sigma_x^2, \phi_x][\boldsymbol{\theta}]$$

with $[\boldsymbol{\theta}]$ the prior distribution for the collection of model parameters.

This procedure enables joint modeling of the covariate and response surfaces despite the spatial misalignment during data collection. From here, sampling of the directional derivatives will occur as described in Chapter 2 which will directly enable post model fitting gradient analyses without further complication.

For the second scenario we assume a continuous response $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ and an areal covariate corresponding to a block average for each areal unit $\{X(B) : B \in \mathcal{B}\}$. For simplicity we assume that the covariate process $X(\mathbf{s})$ follows a spatially varying intercept model, but more complex mean functions could be accommodated analogously. For each observed location $\mathbf{s}$ and each observed areal unit B we have the following model

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s}) \\ X(B) &= |B|^{-1} \int_B X(\mathbf{s}')d\mathbf{s}' \\ X(\mathbf{s}') &= \alpha_0 + w_x(\mathbf{s}') \\ w_y(\mathbf{s}) &\sim GP(0, K(\cdot)) \\ w_x(\mathbf{s}') &\sim GP(0, G(\cdot)) \end{aligned} \tag{5.5}$$

where $K(\cdot) = \rho_y(\cdot; \sigma_y^2, \phi_y)$, $G(\cdot) = \rho_x(\cdot; \sigma_x^2, \phi_x)$ and $|B|$ denotes the area of B. Assuming a mean square differentiable Gaussian process model for the $X(\mathbf{s})$ and $Y(\mathbf{s})$ processes, the joint distribution for $(Y(\mathbf{s}), X(\mathbf{s}), \nabla_Y(\mathbf{s}), \nabla_X(\mathbf{s}), X(B))'$ will be multivariate normal with mean vector $(\mu_Y(\mathbf{s}), \mu_X(\mathbf{s}), \mathbf{0}, \mathbf{0}, \mu_X(B))'$ such that

$$\begin{aligned} \mu_Y(\mathbf{s}) &= \beta_0 + \beta_1 \alpha_0 \\ \mu_X(\mathbf{s}) &= \alpha_0 \\ \mu_X(B) &= |B|^{-1} \int_B \mu_X(\mathbf{s}')d\mathbf{s}' = \alpha_0 \end{aligned}$$

and cross-covariance matrix

$$V(\boldsymbol{\delta}, B_i, B_j) = \begin{pmatrix} \tilde{K}(\boldsymbol{\delta}) & \beta_1 G(\boldsymbol{\delta}) & -\nabla\tilde{K}(\boldsymbol{\delta})' & -\beta_1\nabla G(\boldsymbol{\delta})' & \beta_1 G_{s,B} \\ \beta_1 G(\boldsymbol{\delta}) & G(\boldsymbol{\delta}) & -\beta_1\nabla G(\boldsymbol{\delta})' & -\nabla G(\boldsymbol{\delta})' & G_{s,B} \\ \nabla\tilde{K}(\boldsymbol{\delta}) & \beta_1\nabla G(\boldsymbol{\delta}) & -H_{\tilde{K}}(\boldsymbol{\delta}) & -\beta_1 H_G(\boldsymbol{\delta}) & -\beta_1\nabla G_B \\ \beta_1\nabla G(\boldsymbol{\delta}) & \nabla G(\boldsymbol{\delta}) & -\beta_1 H_G(\boldsymbol{\delta}) & -H_G(\boldsymbol{\delta}) & -\nabla G_B \\ \beta_1 G_{s,B} & G_{s,B} & \beta_1\nabla G_B & \nabla G_B & G_B \end{pmatrix} \tag{5.6}$$

where $\boldsymbol{\delta} = \mathbf{s} - \mathbf{s}'$ and $Cov(Y(\mathbf{s}), Y(\mathbf{s}')) = \tilde{K}(\boldsymbol{\delta}) = K(\boldsymbol{\delta}) + \beta^2 G(\boldsymbol{\delta})$. As in Chapter 2, $\nabla G(\boldsymbol{\delta})$ is the 2×1 gradient vector associated with $G(\boldsymbol{\delta})$, and $H_G(\boldsymbol{\delta})$ is the 2×2 Hessian matrix associated with $G(\boldsymbol{\delta})$. We define the components corresponding to the block averages,

$$(G_B)_{ij} = |B_i|^{-1}|B_j|^{-1}\int_{B_i}\int_{B_j}\rho_x(\mathbf{s}_i - \mathbf{s}_j; \sigma_x^2, \phi_x)d\mathbf{s}_i d\mathbf{s}_j$$

$$(G_{s,B})_{\mathbf{s},j} = |B_j|^{-1}\int_{B_j}\rho_x(\mathbf{s} - \mathbf{s}_j; \sigma_x^2, \phi_x)d\mathbf{s}_j$$

$$(\nabla G_B)_{\mathbf{s},j} = (|B_j|^{-1}\int_{B_j}(\nabla G(\mathbf{s} - \mathbf{s}_j))_1 d\mathbf{s}_j, |B_j|^{-1}\int_{B_j}(\nabla G(\mathbf{s} - \mathbf{s}_j))_2 d\mathbf{s}_j)'$$

I.e., for each location $\mathbf{s}$ and each block B_j, $(\nabla G_B)_{\mathbf{s},j}$ is a 2×1 vector such that element i corresponds to the i^{th} element of $\nabla G(\mathbf{s} - \mathbf{s}_j)$ block averaged over all $\mathbf{s}_j \in B_j$.

In practice the above integrals are approximated using Monte Carlo integration. For each block B_j draw a set of L_j locations $\mathbf{s}_{j1}, \ldots, \mathbf{s}_{jL_j}$ uniformly across the region, with the number of locations L_j adjusted to be appropriate for the block size. The approximations are then

$$(\hat{G}_B)_{ij} = L_i^{-1}L_j^{-1}\Sigma_l\Sigma_m\rho_x(\mathbf{s}_{il} - \mathbf{s}_{jm}; \sigma_x^2, \phi_x) \tag{5.7}$$

$$(\hat{G}_{s,B})_{\mathbf{s},j} = L_j^{-1}\Sigma_l\rho_x(\mathbf{s} - \mathbf{s}_{jl}; \sigma_x^2, \phi_x)$$

$$(\hat{\nabla} G_B)_{\mathbf{s},j} = (L_j^{-1}\Sigma_l(\nabla G(\mathbf{s} - \mathbf{s}_{jl}))_1, L_j^{-1}\Sigma_l(\nabla G(\mathbf{s} - \mathbf{s}_{jl}))_2)'$$

where the "hat" indicates a Monte Carlo integration.

The above joint distribution leads to posterior sampling of the response and covariate gradient vectors analogous to the sampling methodology outlined in Chapter 2. In turn, this enables post model fitting gradient analyses for scenarios when the response has been observed as point-referenced data and the covariate has been observed as block averages across areal units. The gradient analyses would proceed similarly in the case with a block averaged response and a continuous covariate.

For the final scenario we allow both the response and the covariate to be observed as block averages. Say we observe a block average response, like pollution, $\{Y(B) : B \in \mathcal{B}_Y\}$ and a block average covariate $\{X(B) : B \in \mathcal{B}_X\}$ and want to approximate the relative rates of change for the underlying processes. The model follows similarly to (5.5),

$$\begin{aligned}
Y(B) &= |B|^{-1} \int_B Y(\mathbf{s}')d\mathbf{s}' \\
X(B) &= |B|^{-1} \int_B X(\mathbf{s}')d\mathbf{s}' \\
Y(\mathbf{s}) &= \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s}) \\
X(\mathbf{s}) &= \alpha_0 + w_x(\mathbf{s}) \\
w_y(\mathbf{s}) &\sim GP(0, K(\cdot)) \\
w_x(\mathbf{s}) &\sim GP(0, G(\cdot))
\end{aligned}$$

with notation as before. The $Y(\mathbf{s})$ and $X(\mathbf{s})$ processes are assumed to be mean square differentiable in order to satisfy the required smoothness conditions.

The joint process, $(Y(\mathbf{s}), X(\mathbf{s}), \nabla_Y(\mathbf{s}), \nabla_X(\mathbf{s}), Y(B), X(B))'$, will follow a multivariate normal distribution with mean vector $(\mu_Y(\mathbf{s}), \mu_X(\mathbf{s}), \mathbf{0}, \mathbf{0}, \mu_Y(B), \mu_X(B))'$

such that

$$\mu_Y(\mathbf{s}) = \beta_0 + \beta_1\alpha_0$$

$$\mu_X(\mathbf{s}) = \alpha_0$$

$$\mu_Y(B) = |B|^{-1}\int_B \mu_Y(\mathbf{s}')d\mathbf{s}' = \beta_0 + \beta_1\alpha_0$$

$$\mu_X(B) = |B|^{-1}\int_B \mu_X(\mathbf{s}')d\mathbf{s}' = \alpha_0$$

and cross-covariance matrix

$V(\boldsymbol{\delta}, B_i, B_j) =$

$$\begin{pmatrix}
\tilde{K}(\boldsymbol{\delta}) & \beta_1 G(\boldsymbol{\delta}) & -\nabla\tilde{K}(\boldsymbol{\delta})' & -\beta_1\nabla G(\boldsymbol{\delta})' & \tilde{K}_{s,B} & \beta_1 G_{s,B} \\
\beta_1 G(\boldsymbol{\delta}) & G(\boldsymbol{\delta}) & -\beta_1\nabla G(\boldsymbol{\delta})' & -\nabla G(\boldsymbol{\delta})' & \beta_1 G_{s,B} & G_{s,B} \\
\nabla\tilde{K}(\boldsymbol{\delta}) & \beta_1\nabla G(\boldsymbol{\delta}) & -H_{\tilde{K}}(\boldsymbol{\delta}) & -\beta_1 H_G(\boldsymbol{\delta}) & -\nabla\tilde{K}_B & -\beta_1\nabla G_B \\
\beta_1\nabla G(\boldsymbol{\delta}) & \nabla G(\boldsymbol{\delta}) & -\beta_1 H_G(\boldsymbol{\delta}) & -H_G(\boldsymbol{\delta}) & -\beta_1\nabla G_B & -\nabla G_B \\
\tilde{K}_{s,B} & \beta_1 G_{s,B} & \nabla\tilde{K}_B & \beta_1\nabla G_B & \tilde{K}_B & \beta_1 G_B \\
\beta_1 G_{s,B} & G_{s,B} & \beta_1\nabla G_B & \nabla G_B & \beta_1 G_B & G_B
\end{pmatrix}$$

with notation consistent with (5.6). The cross-covariance components related to the block average data can be approximated using Monte Carlo integration as described earlier in (5.7). Using this joint distribution posterior samples of the underlying spatial processes and associated gradients can be drawn for any new location, enabling gradient analyses to be conducted when data is available as block average responses and block average covariates.

5.3 Spatially Varying Coefficient Processes

5.3.1 Introduction

The local directional sensitivity processes developed in previous chapters provide an avenue for learning about the spatial variation in relationships without explicitly modeling those variations. However, in some applications it may be appropriate to explicitly model these relationships through a framework with spatially varying

coefficients (e.g. Fotheringham et al., 2002; Gelfand et al., 2003). Spatially varying coefficient processes (SVCP) provide a fully model-based framework in which the covariate(s) are associated with a different coefficient $\beta(\mathbf{s})$ at each location $\mathbf{s}$, and the $\beta(\mathbf{s})$ process is assumed to follow a Gaussian process model to encourage correlation between proximate locations (Gelfand et al., 2003). Geographically weighted regression (GWR) similarly allows for spatially varying coefficients, but estimates these through spatial weighting in lieu of a spatial model (Fotheringham et al., 2002). Unlike GWR, the model-based nature of SVCP provides a framework in which one can explore the spatially varying relationships in greater detail via the associated local directional sensitivity processes.

Directional derivative processes were previously considered for SVCP models by Majumdar et al. (2006), but here we extend the modeling to include a joint Gaussian process model for a spatial covariate $X(\mathbf{s})$. An outline is provided for post model fitting gradient analyses via the local directional sensitivity process and the spatial angular discrepancy process. While these derivations assume a single covariate, extension to multivariate spatially varying coefficient models follows similarly.

5.3.2 Modeling and Gradient Theory

Following the notation in Gelfand et al. (2003), the spatially varying coefficient model can be written,

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \tilde{\beta}_1(\mathbf{s})X(\mathbf{s}) + w_y(\mathbf{s}) \\ X(\mathbf{s}) &= \alpha_0 + w_x(\mathbf{s}) \\ \tilde{\beta}_1(\mathbf{s}) &= \beta_1 + \beta_1(\mathbf{s}) \end{aligned} \tag{5.8}$$

where $w_y(\mathbf{s})$, $w_x(\mathbf{s})$ and $\beta_1(\mathbf{s})$ are mean-zero Gaussian processes with correlation functions $\rho_y(\sigma_y^2, \phi_y)$, $\rho_x(\sigma_x^2, \phi_x)$ and $\rho_b(1, \phi_b)$. Note that the variance parameter in the correlation for $\beta_1(\mathbf{s})$ is set to 1 to ensure identifiability of the $\beta_1(\mathbf{s})$ and $X(\mathbf{s})$

processes. In this model $\beta_0 + w_y(\mathbf{s})$ and $\alpha_0 + w_x(\mathbf{s})$ can be thought of as spatially varying intercepts centered at β_0 and α_0, and $\tilde{\beta}_1(\mathbf{s})$ as a spatially varying coefficient on $X(\mathbf{s})$ centered at β_1. Potential modifications include the addition of a nugget effect allowing for *iid* error at each location $\mathbf{s}$ and/or the removal of the $w_y(\mathbf{s})$ term leaving a fixed intercept across all locations.

Assuming mean square differentiability of the Gaussian processes, e.g. Matérn covariance functions with smoothness parameter $\nu = 3/2$, the directional derivative process for $Y(\mathbf{s})$ will be a well defined spatial process,

$$D_{\mathbf{u}}Y(\mathbf{s}) = \tilde{\beta}_1(\mathbf{s})D_{\mathbf{u}}X(\mathbf{s}) + X(\mathbf{s})D_{\mathbf{u}}\beta_1(\mathbf{s}) + D_{\mathbf{u}}w_y(\mathbf{s})$$

Variation in the relative rate of change between response $Y(\mathbf{s})$ and covariate $X(\mathbf{s})$ can be assessed across different directional perspectives through gradient analyses using the local directional sensitivity process and the spatial angular discrepancy process,

$$\frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} = \tilde{\beta}_1(\mathbf{s}) + \frac{D_{\mathbf{u}}\beta_1(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})}X(\mathbf{s}) + \frac{D_{\mathbf{u}}w_y(\mathbf{s})}{D_{\mathbf{u}}w_x(\mathbf{s})} \tag{5.9}$$

$$disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_Y(\mathbf{s}))$$

where $\theta_X(\mathbf{s})$ and $\theta_Y(\mathbf{s})$ describe the directions of maximum gradient for each of the variables at location $\mathbf{s}$.

Interestingly, the expression for the local directional sensitivity process in (5.9) has the appearance of a regression on the covariate $X(\mathbf{s})$. The intercept component $\tilde{\beta}_1(\mathbf{s})$ corresponds to the local response-covariate relationship dictated by the model, and the Cauchy ratio $D_{\mathbf{u}}w_y(\mathbf{s})/D_{\mathbf{u}}w_x(\mathbf{s})$ serves as an error term centered at zero as before. It can be shown that the remaining term, $D_{\mathbf{u}}\beta_1(\mathbf{s})/D_{\mathbf{u}}w_x(\mathbf{s})X(\mathbf{s})$, will also be centered at zero so that the overall centering for the local directional sensitivity will be $\tilde{\beta}_1(\mathbf{s})$. However, in general this "regression" term indicates that the response-covariate relationship in direction $\mathbf{u}$ may depend on the covariate value at that location.

In summary, gradient analyses applied to SVCP models inform on the variation in local variable relationships across directional perspectives. Similar to the non-spatially varying coefficient models discussed in earlier chapters, the local directional sensitivity process will be centered at the model parameter defining the response-covariate relationship. In this case the model parameter already provides a local description for the relationship, and the sensitivities will primarily inform on the impact of directional changes. In addition to the centering term the sensitivity consists of two error terms. One corresponds to zero-centered Cauchy noise, and the other allows for positive or negative adjustments to the relative rates of change based on the covariate value and the relative rates of change of the coefficient and covariate at location $\mathbf{s}$.

5.4 Predictive Process Model

5.4.1 Introduction

As technologies such as Global Positioning Systems (GPS) have improved the accuracy of geocoded locations, spatial data has become increasingly popular with an ever increasing number of observed variables, locations and time points. Bayesian hierarchical spatial modeling generally provides a flexible framework for accommodating spatial data, but computationally suffers in large dimensions due to expensive matrix decompositions. In recent years several approaches have been proposed to accommodate large spatial data. A common tactic is to replace the full spatial process with an approximation that lies in a lower dimensional subspace such as basis functions (e.g. Wikle and Cressie, 1999; Paciorek, 2007), kernel convolutions (e.g. Higdon, 2002), or moving averages (e.g. Ver Hoef et al., 2004). Another approach is to approximate the full likelihood by working in the spectral domain of the spatial process thus avoiding the matrix computations (e.g. Stein, 1999; Fuentes, 2007; Paciorek, 2007), or paritioning the space into independent contiguous blocks (e.g. Kim

et al., 2005; Konomi et al., 2013; Stein, 2013).

The gradient analyses proposed in previous chapters rely on posterior samples of the parameters in the parent spatial model. However, in high dimensions it may be too computationally challenging to fit the full parent model, with researchers instead turning to a dimension reduction technique. Despite the need to approximate the model, local variation in the response-covariate relationships may still be of interest and can be examined through post model fitting gradient analyses analogous to those derived for the small data setting. To illustrate how post model fitting gradient analyses can apply to large spatial data we focus on the dimension reduction technique known as predictive processes (Banerjee et al., 2008; Finley et al., 2009). Predictive process models were proposed by Banerjee et al. (2008) where the predictive processes are induced by a full spatial process model, projecting the process realizations onto a lower dimensional space based on a set of spatial knots. The modified predictive process was later proposed to address a positive bias in the non-spatial error term in the original methodology, providing an algorithm for optimal placement of the knots Finley et al. (2009).

5.4.2 Modeling and Gradient Theory

As with the small data setting, we imagine a scenario where both the response and the covariate of interest are assumed to follow a spatial process model,

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \beta_1 X(\mathbf{s}) + w_y(\mathbf{s}) \\ X(\mathbf{s}) &= \alpha_0 + w_x(\mathbf{s}) \\ w_y(\mathbf{s}) &\sim GP(0, C_y(\theta_y)) \\ w_x(\mathbf{s}) &\sim GP(0, C_x(\theta_x)) \end{aligned} \tag{5.10}$$

where $w_y(\mathbf{s})$ and $w_x(\mathbf{s})$ are independent Gaussian processes with covariance functions $C_y(\theta_y)$ and $C_x(\theta_x)$. Due to the large dimension of the data the full Gaussian process

model is assumed to be computationally infeasible, so we turn to the induced joint predictive process model to relate the response $Y(\mathbf{s})$ and covariate $X(\mathbf{s})$. Following the notation from Banerjee et al. (2008), this gives the following model

$$\begin{aligned} Y(\mathbf{s}) &= \beta_0 + \beta_1 X(\mathbf{s}) + \tilde{w}_y(\mathbf{s}) \\ X(\mathbf{s}) &= \alpha_0 + \tilde{w}_x(\mathbf{s}) \\ \tilde{w}_y(\mathbf{s}) &= \mathbf{c}'_{\mathbf{y}}(\mathbf{s})\mathbf{C}^{*-\mathbf{1}}_{\mathbf{y}}(\theta)\mathbf{w}^*_{\mathbf{y}} \\ \tilde{w}_x(\mathbf{s}) &= \mathbf{c}'_{\mathbf{x}}(\mathbf{s})\mathbf{C}^{*-\mathbf{1}}_{\mathbf{x}}(\theta)\mathbf{w}^*_{\mathbf{x}} \end{aligned} \tag{5.11}$$

where $\mathcal{S}^* = \{\mathbf{s}^*_1, \ldots, \mathbf{s}^*_m\}$ is a set of spatial knot locations, $\mathbf{w}^* = [w(\mathbf{s}^*_i)]^m_{i=1} \sim N(0, \mathbf{C}^*(\theta))$ is a realization from the Gaussian process in (5.10) at the set of knot locations, $\mathbf{C}^*(\theta) = [C(\mathbf{s}^*_i, \mathbf{s}^*_j|\theta)]^m_{i,j=1}$ and $\mathbf{c}(\mathbf{s}) = [C(\mathbf{s}, \mathbf{s}^*_j|\theta)]^m_{j=1}$. Note, there is no pure or measurement error $\epsilon(\mathbf{s})$ term being used; the motivational data consists of temperature and elevation observations, both of which can be reasonably modeled with purely spatial error. For data that do require a pure error term, the gradient analyses would be applied to the mean surface and would be carried out in the same manner outlined here.

When computing directional derivatives associated with predictive process models, we need only consider the derivative associated with the covariance function in the $\mathbf{c}'(\mathbf{s})$ matrix since the remaining terms depend only on the parameters and the spatial knot locations $\mathcal{S}^*$. For example, we can write the directional derivative for $Y(\mathbf{s})$ and $X(\mathbf{s})$ in (5.11) as

$$\begin{aligned} D_{\mathbf{u}}Y(\mathbf{s}) &= \beta_1 D_{\mathbf{u}}X(\mathbf{s}) + D_{\mathbf{u}}\tilde{w}_y(\mathbf{s}) \\ D_{\mathbf{u}}X(\mathbf{s}) &= D_{\mathbf{u}}\tilde{w}_x(\mathbf{s}) \\ D_{\mathbf{u}}\tilde{w}_y(\mathbf{s}) &= D_{\mathbf{u}}\mathbf{c}_y(\mathbf{s})\mathbf{C}^{*-\mathbf{1}}_{\mathbf{y}}(\theta)\mathbf{w}^*_{\mathbf{y}} \\ D_{\mathbf{u}}\tilde{w}_x(\mathbf{s}) &= D_{\mathbf{u}}\mathbf{c}_x(\mathbf{s})\mathbf{C}^{*-\mathbf{1}}_{\mathbf{x}}(\theta)\mathbf{w}^*_{\mathbf{x}} \end{aligned} \tag{5.12}$$

where $D_{\mathbf{u}}\mathbf{c}(\mathbf{s}) = [D_{\mathbf{u}}c(\mathbf{s}, \mathbf{s}^*_i)]^m_{i=1}$ is a matrix of directional derivatives associated with

the covariance function $c(\cdot)$. For a covariance function $c(\mathbf{s}, \mathbf{s}_i^*)$, the directional derivative in the direction $\mathbf{u}$ will be calculated in the standard way

$$\begin{aligned} D_{\mathbf{u}}c(\mathbf{s}, \mathbf{s}_i^*) &= \lim_{h \to 0} \frac{c(\mathbf{s} + h\mathbf{u}, \mathbf{s}_i^*) - c(\mathbf{s}, \mathbf{s}_i^*)}{h} \\ &= [\frac{d}{dh} c(\mathbf{s} + h\mathbf{u}, \mathbf{s}_i^*)]_{h=0} \end{aligned}$$

where $\mathbf{s}$ is the location of interest and $\mathbf{s}_i^*$ is a knot in the predictive process. In our examples we consider the Matérn covariance function, $c(\mathbf{s}, \mathbf{s}_i^*) = \sigma^2(1 + \phi||\mathbf{s} - \mathbf{s}_i^*||)\exp(-\phi||\mathbf{s} - \mathbf{s}_i^*||)$, with associated directional derivative $D_{\mathbf{u}}c(\mathbf{s}, \mathbf{s}_i^*) = -\sigma^2\phi^2(\mathbf{s} - \mathbf{s}_i^*)'\mathbf{u}\exp(-\phi||\mathbf{s} - \mathbf{s}_i^*||)$.

Post model fitting, we can compute samples of the directional derivatives associated with $Y(\mathbf{s})$ using (5.12) and the posterior samples of the model parameters, and similarly for $X(\mathbf{s})$. To compute the local directional sensitivity process we have

$$\begin{aligned} \frac{D_{\mathbf{u}}Y(\mathbf{s})}{D_{\mathbf{u}}X(\mathbf{s})} &= \beta_1 + \frac{D_{\mathbf{u}}\tilde{w}_y(\mathbf{s})}{D_{\mathbf{u}}\tilde{w}_x(\mathbf{s})} \\ &= \beta_1 + \frac{D_{\mathbf{u}}\mathbf{c}_{\mathbf{y}}'(\mathbf{s})\mathbf{C}_{\mathbf{y}}^{*-\mathbf{1}}(\theta)\mathbf{w}_{\mathbf{y}}^*}{D_{\mathbf{u}}\mathbf{c}_{\mathbf{x}}'(\mathbf{s})\mathbf{C}_{\mathbf{x}}^{*-\mathbf{1}}(\theta)\mathbf{w}_{\mathbf{x}}^*} \end{aligned} \tag{5.13}$$

As before, this process is centered at the global relationship parameter β_1 and follows a Cauchy distribution. The Cauchy distribution can be easily seen since this is again a ratio of independent normal random variables, with scale dependent on the covariance functions $c_y(\cdot)$ and $c_x(\cdot)$ and their derivatives in the direction $\mathbf{u}$.

In addition, the spatial angular discrepancy process can be computed in a fashion analogous to the standard Gaussian process models. The direction of maximum gradient will occur in the direction described by the unit vector $\nabla_Y(\mathbf{s})/||\nabla_Y(\mathbf{s})||$, which corresponds to the angle $\theta_Y(\mathbf{s})$. The spatial angular discrepancy can then be computed as $disc(\mathbf{s}) = 1 - cos(\theta_X(\mathbf{s}) - \theta_Y(\mathbf{s}))$.

5.5 Total Derivative

5.5.1 Introduction

In many applications one may be interested in two spatial processes where the linear relationship assumed in previous chapters will be unrealistic. Instead, the researcher can describe their relationship using some function of the two processes that is assumed to exhibit spatial variation. An example would be circular behavior between $X(\mathbf{s})$ and $Y(\mathbf{s})$,

$$X(\mathbf{s})^2 + Y(\mathbf{s})^2 = r^2 + w_g(\mathbf{s}) \tag{5.14}$$
$$w_g(\mathbf{s}) \sim GP(0, \rho_g(\sigma_g^2, \phi_g))$$

where the circular relationship exhibits spatial error described by $w_y(\mathbf{s})$ and one of the processes is assumed to have a standard linear form

$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$
$$w_x(\mathbf{s}) \sim GP(0, \rho_x(\sigma_x^2, \phi_x))$$

In this way we model $X(\mathbf{s})$ as a covariate and learn about $Y(\mathbf{s})$ through its known (non-linear) relationship with $X(\mathbf{s})$. To learn about the relative rates of change of the two variables, we then rely on the spatial equivalent of the *total derivative* used in standard calculus. Examination of the relative rates of change will provide insight into their relative behaviors at any spatial location, after having constrained the processes to conform to the non-linear relationship.

By outlining gradient theory for spatial processes with known non-linear relationships, we open the spatial gradient analyses to a broad range of potential applications.

5.5.2 Modeling and Gradient Theory

Assume that the relationship between $X(\mathbf{s})$ and $Y(\mathbf{s})$ is described by a differentiable function $g(\cdot)$ and the covariate $X(\mathbf{s})$ has a linear form,

$$g(X(\mathbf{s}), Y(\mathbf{s})) = w_g(\mathbf{s})$$
$$X(\mathbf{s}) = \alpha_0 + w_x(\mathbf{s})$$

Then the associated directional derivatives associated with the two processes can be written

$$\frac{\partial g}{\partial X(\mathbf{s})} D_{\mathbf{u}} X(\mathbf{s}) + \frac{\partial g}{\partial Y(\mathbf{s})} D_{\mathbf{u}} Y(\mathbf{s}) = D_{\mathbf{u}} w_g(\mathbf{s})$$
$$D_{\mathbf{u}} X(\mathbf{s}) = D_{\mathbf{u}} w_x(\mathbf{s})$$

Solving for $D_{\mathbf{u}} Y(\mathbf{s})$ in the above expression allows us to learn about its directional rates of change. The local directional sensitivity process considers these rates of change relative to those for the $X(\mathbf{s})$ process,

$$\frac{D_{\mathbf{u}} Y(\mathbf{s})}{D_{\mathbf{u}} X(\mathbf{s})} = \frac{1}{\partial g / \partial Y(\mathbf{s})} \frac{D_{\mathbf{u}} w_g(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} - \frac{\partial g / \partial X(\mathbf{s})}{\partial g / \partial Y(\mathbf{s})}$$

The above sensitivity involves two terms. The first term corresponds to the relative rates of change in the direction $\mathbf{u}$ for the two spatial processes, $w_g(\mathbf{s})$ and $w_x(\mathbf{s})$, weighted by one over the partial derivative with respect to $Y(\mathbf{s})$. The second term is the negative of the ratio of partial derivatives, taking into account the relative importance of the two variables in the function $g(\cdot)$. For the circular example in (5.14) the local directional sensitivity will be

$$\frac{D_{\mathbf{u}} Y(\mathbf{s})}{D_{\mathbf{u}} X(\mathbf{s})} = \frac{1}{2Y(\mathbf{s})} \frac{D_{\mathbf{u}} w_g(\mathbf{s})}{D_{\mathbf{u}} w_x(\mathbf{s})} - \frac{X(\mathbf{s})}{Y(\mathbf{s})}$$

In addition to allowing for non-linear functions of the response and covariate, the total derivative allows for gradient computation when the response depends

on differentiable functions of multiple covariates. For example, recall the interaction model discussed in Chapter 4. Assuming $Y(\mathbf{s}) = \beta_0 + \beta_1 X_1(\mathbf{s}) + \beta_2 X_2(\mathbf{s}) + \beta_3 X_1(\mathbf{s}) X_2(\mathbf{s}) + w_y(\mathbf{s})$, we have a differentiable function of the $X_1(\mathbf{s})$ and $X_2(\mathbf{s})$ processes. The directional derivative for $Y(\mathbf{s})$ is then $D_{\mathbf{u}} Y(\mathbf{s}) = \beta_1 D_{\mathbf{u}} X_1(\mathbf{s}) + \beta_2 D_{\mathbf{u}} X_2(\mathbf{s}) + \beta_3 X_1(\mathbf{s}) D_{\mathbf{u}} X_2(\mathbf{s}) + \beta_3 X_2(\mathbf{s}) D_{\mathbf{u}} X_1(\mathbf{s})$. Other non-linear functions of the covariates can be similarly accommodated through the total derivative.

Appendix A

Bivariate Gradient Theory

A.1 Spatial Cauchy Process

We show that $D_{\mathbf{u}}w_y(\mathbf{s})/D_{\mathbf{u}}w_x(\mathbf{s})$ is a well defined spatial stochastic process.

First consider two locations $\mathbf{s}$ and $\mathbf{s}'$. For notational convenience, let $D_{\mathbf{u}}w_y(\mathbf{s}) = n_1$, $D_{\mathbf{u}}w_y(\mathbf{s}') = n_2$, $D_{\mathbf{u}}w_x(\mathbf{s}) = m_1$, and $D_{\mathbf{u}}w_x(\mathbf{s}') = m_2$. We can write the joint distribution of the ratios in terms of the joint distributions for the two Gaussian

processes:

$$
\begin{aligned}
P(\frac{n_1}{m_1} < r_1,\ \frac{n_2}{m_2} < r_2) &= P(n_1 < r_1 m_1,\ n_2 < r_2 m_2,\ m_1 > 0,\ m_2 > 0) \\
&+ P(n_1 < r_1 m_1,\ n_2 > r_2 m_2,\ m_1 > 0,\ m_2 < 0) \\
&+ P(n_1 > r_1 m_1\ , n_2 < r_2 m_2,\ m_1 < 0,\ m_2 > 0) \\
&+ P(n_1 > r_1 m_1,\ n_2 > r_2 m_2,\ m_1 < 0,\ m_2 < 0) \\
&= \int_0^\infty \int_0^\infty \int_{-\infty}^{r_1 m_1} \int_{-\infty}^{r_2 m_2} f_K(n_1, n_2) f_G(m_1, m_2) dn_2 dn_1 dm_2 dm_1 \\
&+ \int_0^\infty \int_{-\infty}^0 \int_{-\infty}^{r_1 m_1} \int_{r_2 m_2}^\infty f_K(n_1, n_2) f_G(m_1, m_2) dn_2 dn_1 dm_2 dm_1 \\
&+ \int_{-\infty}^0 \int_0^\infty \int_{r_1 m_1}^\infty \int_{-\infty}^{r_2 m_2} f_K(n_1, n_2) f_G(m_1, m_2) dn_2 dn_1 dm_2 dm_1 \\
&+ \int_{-\infty}^0 \int_{-\infty}^0 \int_{r_1 m_1}^\infty \int_{r_2 m_2}^\infty f_K(n_1, n_2) f_G(m_1, m_2) dn_2 dn_1 dm_2 dm_1
\end{aligned}
$$

The above joint distribution depends only on integrals of multivariate normal densities, so consistency across permutations of the labels is clearly satisfied.

At a single location the directional derivative ratio is a ratio of two independent normal random variables. As such, it will have a Cauchy distribution with a scale parameter dependent on the parameters in the covariance functions $K(\cdot)$ and $G(\cdot)$, namely $SD(n_1)/SD(m_1)$. Next we show that marginalizing the bivariate distribution

over r_2 will reduce to the known distribution for the univariate case:

$$
\begin{aligned}
P(\frac{n_1}{m_1} < r_1) &= \int_{-\infty}^{\infty} P(\frac{n_1}{m_1} < r_1,\ \frac{n_2}{m_2} < r_2) dr_2 \\
&= \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{r_1 m_1}\int_{-\infty}^{r_2 m_2} f_K(n_1,n_2) f_G(m_1,m_2) dn_2 dn_1 dm_2 dm_1 dr_2 \\
&+ \int_{-\infty}^{\infty}\int_{-\infty}^{0}\int_{0}^{\infty}\int_{-\infty}^{r_1 m_1}\int_{r_2 m_2}^{\infty} f_K(n_1,n_2) f_G(m_1,m_2) dn_2 dn_1 dm_2 dm_1 dr_2 \\
&+ \int_{-\infty}^{\infty}\int_{-\infty}^{0}\int_{0}^{\infty}\int_{r_1 m_1}^{\infty}\int_{-\infty}^{r_2 m_2} f_K(n_1,n_2) f_G(m_1,m_2) dn_2 dn_1 dm_2 dm_1 dr_2 \\
&+ \int_{-\infty}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{0}\int_{r_1 m_1}^{\infty}\int_{r_2 m_2}^{\infty} f_K(n_1,n_2) f_G(m_1,m_2) dn_2 dn_1 dm_2 dm_1 dr_2 \\
&= \int_{0}^{\infty}\int_{0}^{\infty}\int_{-\infty}^{r_1 m_1} f_K(n_1) f_G(m_1,m_2) dn_1 dm_2 dm_1 \\
&+ \int_{0}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{r_1 m_1} f_K(n_1) f_G(m_1,m_2) dn_1 dm_2 dm_1 \\
&+ \int_{-\infty}^{0}\int_{0}^{\infty}\int_{r_1 m_1}^{\infty} f_K(n_1) f_G(m_1,m_2) dn_1 dm_2 dm_1 \\
&+ \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{r_1 m_1}^{\infty} f_K(n_1) f_G(m_1,m_2) dn_1 dm_2 dm_1 \\
&= \int_{0}^{\infty}\int_{-\infty}^{r_1 m_1} f_k(n_1) f_G(m_1) dn_1 dm_1 + \int_{-\infty}^{0}\int_{r_1 m_1}^{\infty} f_k(n_1) f_G(m_1) dn_1 dm_1
\end{aligned}
$$

We can then rewrite this as:

$$
\begin{aligned}
P(\frac{n_1}{m_1} < r_1) &= \int_{0}^{\infty} F_K(r_1 m_1) f_G(m_1) dm_1 + \int_{-\infty}^{0} F_K(-r_1 m_1) f_G(m_1) dm_1 \\
&= 2\int_{0}^{\infty} F_K(r_1 m_1) f_G(m_1) dm_1
\end{aligned}
$$

The associated density will then be $2\int_0^\infty m_1 f_K(r_1 m_1) f_G(m_1) dm_1$, which was shown in Andrews and Mallows (1974) to be a Cauchy distribution. In this case we will have a Cauchy with scale parameter $SD(n_1)/SD(m_1)$, as desired.

The marginalization is straight forward for larger dimensions with marginalization over the two normal densities occurring in a similar way.

A.2 Bivariate Angular Density

We are interested in the marginal behavior of $f(\theta_X(\mathbf{s}), \theta_Y(\mathbf{s}))$ at a location $\mathbf{s}$ where $\tan(\theta_X(\mathbf{s})) = D_{(0,1)}X(\mathbf{s})/D_{(1,0)}X(\mathbf{s})$, $\tan(\theta_Y(\mathbf{s})) = D_{(0,1)}Y(\mathbf{s})/D_{(1,0)}Y(\mathbf{s})$, and $(\nabla_X(\mathbf{s}), \nabla_Y(\mathbf{s}))' = (D_{(1,0)}Y(\mathbf{s}), D_{(0,1)}Y(\mathbf{s}), D_{(1,0)}X(\mathbf{s}), D_{(0,1)}X(\mathbf{s}))'$ are normally distributed with the covariance structure provided in Equation 1.

After converting to polar coordinates, the change of variables formula gives the following integral, where $g(\cdot)$ is the multivariate normal density associated with $(\nabla_Y(\mathbf{s}), \nabla_X(\mathbf{s}))'$. For clarity we suppress the index $(\mathbf{s})$:

$$
\begin{aligned}
f(\theta_X, \theta_Y) = &\int_0^\infty \int_0^\infty g(r_y \cos\theta_Y, r_y \sin\theta_Y, r_x \cos\theta_X, r_x \sin\theta_X) r_y r_x dr_y dr_x \\
= &\int_0^\infty r_x \frac{1}{\sqrt{(2\pi)^3|\Sigma|}} \exp(-\frac{1}{2}(ac - \tilde{A}^2) r_x^2) \\
&\times \left(\int_0^\infty r_y \frac{\sqrt{a}}{\sqrt{2\pi}} \exp(-\frac{a}{2}(r_y - \beta r_x \cos\theta)^2) dr_y \right) dr_x \\
= &\frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \\
&\times \int_0^\infty \frac{1}{\sqrt{2\pi}} \left(\frac{\Phi(\tilde{A} r_x)\tilde{A} r_x/\sqrt{a}}{\sqrt{a}} + \frac{\phi(\tilde{A} r_x)}{a} \right) r_x \exp(-\frac{1}{2}(ac - \tilde{A}^2) r_x^2) dr_x \\
= &\frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \Big(\frac{1}{\sqrt{2\pi a}} \frac{\tilde{A}}{\sqrt{a}} \int_0^\infty r_x \Phi(\tilde{A} r_x) r_x \exp(-\frac{1}{2}(ac - \tilde{A}^2) r_x^2) dr_x \\
&+ \int_0^\infty r_x \frac{1}{a\sqrt{2\pi}} \phi(\tilde{A} r_x) \exp(-\frac{1}{2}(ac - \tilde{A}^2) r_x^2) dr_x \Big)
\end{aligned}
$$

$$= \begin{cases} d\left(\frac{1}{\sqrt{2\pi a}}\frac{\tilde{A}}{\sqrt{a}}\left(\frac{\phi(0)\tilde{A}}{ac(ac-\tilde{A}^2)}+\frac{\sqrt{2\pi}}{(ac-\tilde{A}^2)^{3/2}}L(0,0,\sqrt{\frac{\tilde{A}^2}{ac}})\right)+\frac{\phi(0)}{a^2c\sqrt{2\pi}}\right) & \text{, if } \tilde{A}>0 \\ d\left(\frac{1}{\sqrt{2\pi a}}\frac{\tilde{A}}{\sqrt{a}}\left(\frac{\phi(0)\tilde{A}}{ac(ac-\tilde{A}^2)}+\frac{\sqrt{2\pi}}{(ac-\tilde{A}^2)^{3/2}}(0.5-L(0,0,\sqrt{\frac{\tilde{A}^2}{ac}}))\right)+\frac{\phi(0)}{a^2c\sqrt{2\pi}}\right) & \text{, if } \tilde{A}<0 \end{cases}$$

where $d = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}}$, $|\Sigma| = (\sigma_x^2\phi_x^2)^2(\sigma_y^2\phi_y^2)^2$, $a = 1/(\sigma_y^2\phi_y^2)$, $c = (\sigma_y^2\phi_y^2+\beta^2\phi_x^2\sigma_x^2)/(\sigma_x^2\phi_x^2)$, $\tilde{A} = \sqrt{a}\beta\cos(\theta_X-\theta_Y)$, and $L(0,0,\rho)$ is the zero mean bivariate normal cdf with correlation ρ and standard deviations equal to 1 evaluated at $(0,0)'$.

Several integration steps were treated as the expected value of a truncated normal distribution. An additional integration step required integration by three parts, then computation of the probability that the sum of a normal and a truncated normal are less than some constant. This probability was available in closed form in the query by Lipow et al. (1964) and is written in terms of the bivariate normal cdf. Somewhat simplified, this gives the density provided in the text:

$$f(\theta_X,\theta_Y) = \begin{cases} C\left(\frac{\tilde{A}^2\phi(0)}{ac(ac-\tilde{A}^2)}+\frac{\sqrt{2\pi}\tilde{A}}{(ac-\tilde{A}^2)^{3/2}}L(0,0,\sqrt{\frac{\tilde{A}^2}{ac}})+\frac{\phi(0)}{ac}\right) & \text{, if } \tilde{A}>0 \\ C\left(\frac{\tilde{A}^2\phi(0)}{ac(ac-\tilde{A}^2)}+\frac{\sqrt{2\pi}\tilde{A}}{(ac-\tilde{A}^2)^{3/2}}\left(0.5-L(0,0,\sqrt{\frac{\tilde{A}^2}{ac}})\right)+\frac{\phi(0)}{ac}\right) & \text{, if } \tilde{A}<0 \end{cases}$$

where $C = \frac{1}{a(2\pi)^{3/2}\sqrt{|\Sigma|}}$.

Bibliography

Albert, J. H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

Andrews, D. F. and Mallows, C. L. (1974), "Scale mixtures of normal distributions," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 99–102.

Banerjee, S. and Gelfand, A. (2003), "On smoothness properties of spatial processes," *Journal of Multivariate Analysis*, 84, 85–100.

Banerjee, S. and Gelfand, A. E. (2006), "Bayesian wombling: Curvilinear gradient assessment under spatial process models," *Journal of the American Statistical Association*, 101, 1487–1501.

Banerjee, S., Gelfand, A. E., and Sirmans, C. (2003), "Directional rates of change under spatial process models," *Journal of the American Statistical Association*, 98, 946–954.

Banerjee, S., Gelfand, A. E., and Carlin, B. P. (2004), *Hierarchical modeling and analysis for spatial data*, CRC Press.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian predictive process models for large spatial data sets," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 825–848.

Baribault, T. W., Kobe, R. K., and Finley, A. O. (2012), "Tropical tree growth is correlated with soil phosphorus, potassium, and calcium, though not for legumes," *Ecological Monographs*, 82, 189–203.

Berrett, C. and Calder, C. A. (2012), "Data augmentation strategies for the Bayesian spatial probit regression model," *Computational Statistics & Data Analysis*, 56, 478–490.

Canham, C. D. and Thomas, R. Q. (2010), "Frequency, not relative abundance, of temperate tree species varies along climate gradients in eastern North America," *Ecology*, 91, 3433–3440.

Clark, J. S. (2005), "Why environmental scientists are becoming Bayesians," *Ecology letters*, 8, 2–14.

Clark, J. S. and Gelfand, A. E. (2006), "A future for models and data in environmental science," *Trends in Ecology & Evolution*, 21, 375–380.

Cressie, N. and Wikle, C. K. (2011), *Statistics for spatio-temporal data*, John Wiley & Sons.

Cressie, N., Calder, C. A., Clark, J. S., Hoef, J. M. V., and Wikle, C. K. (2009), "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling," *Ecological Applications*, 19, 553–570.

Finley, A. O., Banerjee, S., and Carlin, B. P. (2007), "spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models," *Journal of Statistical Software*, 19, 1.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009), "Improving the performance of predictive process modeling for large datasets," *Computational statistics & data analysis*, 53, 2873–2884.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002), *Geographically weighted regression*, Wiley New York.

Fuentes, M. (2007), "Approximate likelihood for large irregularly spaced spatial data," *Journal of the American Statistical Association*, 102, 321–331.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003), "Spatial modeling with spatially varying coefficient processes," *Journal of the American Statistical Association*, 98, 387–396.

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005), "Spatial process modelling for univariate and multivariate dynamic spatial data," *Environmetrics*, 16, 465–479.

Gelfand, A. E., Silander, J. A., Wu, S., Latimer, A., Lewis, P. O., Rebelo, A. G., and Holder, M. (2006), "Explaining species distribution patterns through hierarchical modeling," *Bayesian Analysis*, 1, 41–92.

Gotway, C. A. and Young, L. J. (2002), "Combining incompatible spatial data," *Journal of the American Statistical Association*, 97, 632–648.

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2009), "Measurement error caused by spatial misalignment in environmental epidemiology," *Biostatistics*, 10, 258–274.

Guindani, M. and Gelfand, A. E. (2006), "Smoothness properties and gradient analysis under spatial Dirichlet process models," *Methodology and Computing in Applied Probability*, 8, 159–189.

Heagerty, P. J. and Lele, S. R. (1998), "A composite likelihood approach to binary spatial data," *Journal of the American Statistical Association*, 93, 1099–1111.

Higdon, D. (2002), "Space and space-time modeling using process convolutions," in *Quantitative methods for current environmental issues*, pp. 37–56, Springer.

Jammalamadaka, S. R. and Sengupta, A. (2001), *Topics in circular statistics*, vol. 5, World Scientific.

Kent, J. T. (1989), "Continuity properties for random fields," *The Annals of Probability*, 17, 1432–1440.

Kim, H.-M., Mallick, B. K., and Holmes, C. (2005), "Analyzing nonstationary spatial data using piecewise Gaussian processes," *Journal of the American Statistical Association*, 100, 653–668.

Konomi, B. A., Sang, H., and Mallick, B. K. (2013), "Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations," *Journal of Computational and Graphical Statistics*.

Lamont, B. B., Groom, P. K., and Cowling, R. (2002), "High leaf mass per area of related species assemblages may reflect low rainfall and carbon isotope discrimination rather than low phosphorus and nitrogen concentrations," *Functional Ecology*, 16, 403–412.

Lipow, M., Mantel, N., and Wilkinson, J. (1964), "Query 2: The Sum of Values from a Normal and a Truncated Normal Distribution (Continued)," *Technometrics*, 6, 469–471.

Majumdar, A., Munneke, H. J., Gelfand, A. E., Banerjee, S., and Sirmans, C. (2006), "Gradients in spatial response surfaces with application to urban land values," *Journal of Business & Economic Statistics*, 24.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998), "Log gaussian cox processes," *Scandinavian Journal of Statistics*, 25, 451–482.

Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (2000), "Fully model-based approaches for spatially misaligned data," *Journal of the American Statistical Association*, 95, 877–887.

Murray, I. and Adams, R. P. (2010), "Slice sampling covariance hyperparameters of latent Gaussian models," *arXiv preprint arXiv:1006.0868*.

Murray, I., Adams, R. P., and MacKay, D. J. (2009), "Elliptical slice sampling," *arXiv preprint arXiv:1001.0175*.

Paciorek, C. J. (2007), "Computational techniques for spatial logistic regression with large data sets," *Computational statistics & data analysis*, 51, 3631–3653.

Peng, R. D. and Bell, M. L. (2010), "Spatial misalignment in time series studies of air pollution and health data," *Biostatistics*, 11, 720–740.

Poorter, H., Niinemets, Ü., Poorter, L., Wright, I. J., and Villar, R. (2009), "Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis," *New Phytologist*, 182, 565–588.

Quick, H., Banerjee, S., Carlin, B. P., et al. (2013), "Modeling temporal gradients in regionally aggregated California asthma hospitalization data," *The Annals of Applied Statistics*, 7, 154–176.

Rebelo, A. (2002), "The Protea Atlas Project," Technical report. Accessed: 2002-05-12, `http://protea.worldonline.co.za/default.htm`.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011a), "Bayesian spatial quantile regression," *Journal of the American Statistical Association*, 106.

Reich, B. J., Eidsvik, J., Guindani, M., Nail, A. J., Schmidt, A. M., et al. (2011b), "A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration," *The Annals of Applied Statistics*, 5, 2425–2447.

Royle, J. A. and Berliner, L. M. (1999), "A hierarchical approach to multivariate spatial modeling and prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, pp. 29–56.

Royle, J. A., Chandler, R. B., Gazenski, K. D., and Graves, T. A. (2013), "Spatial capture-recapture models for jointly estimating population density and landscape connectivity," *Ecology*, 94, 287–294.

Schultze, R. (1997), "South African Atlas of Agrohyrdology and Climatology." Technical report, Report TT82/96, Water Research Commission, Pretoria, South Africa.

Stein, M. L. (1999), *Interpolation of spatial data: some theory for kriging*, Springer.

Stein, M. L. (2013), "Limitations on low rank approximations for covariance matrices of spatial data," *Spatial Statistics*.

Takhtajan, A., Crovello, T. J., and Cronquist, A. (1986), *Floristic regions of the world*, University of California press Berkeley.

Thomas, C. D. (2010), "Climate, climate change and range boundaries," *Diversity and Distributions*, 16, 488–495.

Thuiller, W., Lavorel, S., Midgley, G., Lavergne, S., and Rebelo, T. (2004), "Relating plant traits and species distributions along bioclimatic gradients for 88 Leucadendron taxa," *Ecology*, 85, 1688–1699.

Tomovic, R. and Vukobratovic, M. (1972), *General Sensitivity Theory. Number 35 in Modern analytic and computational methods in science and mathematics*, Elsevier, New York.

Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004), "Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, 13, 265–282.

Wang, F. (2013), "Space and Space-Time Modeling of Directional Data," PhD dissertation, Duke University.

Wheeler, D. C. and Calder, C. A. (2007), "An assessment of coefficient accuracy in linear regression models with spatially varying coefficients," *Journal of Geographical Systems*, 9, 145–166.

Wheeler, D. C. and Waller, L. A. (2009), "Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests," *Journal of Geographical Systems*, 11, 1–22.

Wikle, C. K. and Cressie, N. (1999), "A dimension-reduced approach to space-time Kalman filtering," *Biometrika*, 86, 815–829.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998), "Hierarchical Bayesian space-time models," *Environmental and Ecological Statistics*, 5, 117–154.

Biography

Maria Antonia Terres was born in Nashua, N.H. on April 24, 1988. She graduated with a Bachelor of Arts in Biology and Mathematics from Bard College at Simon's Rock in 2008 and a Master of Arts in Statistics from Columbia University in 2010. In 2012 she earned a Master of Science in Statistical Science *en route* to the Ph.D. She graduated with a Doctor of Philosophy under the supervision of Prof. Alan Gelfand in 2014 from Duke University. As a graduate student at Duke she was awarded the James B. Duke Fellowship and the 2014 Dean's Award for Excellence in Teaching. After completion of her Ph.D., she will be joining the Department of Statistics at North Carolina State University as a postdoctoral research scholar under the supervision of Prof. Montserrat Fuentes.